



DATA DE CONFIANCE

LE PARTAGE DES DONNÉES,
CLÉ DE NOTRE AUTONOMIE
STRATÉGIQUE

Olivier Dion et Arno Pons

THINK-TANK
**DIGITAL
NEW DEAL**

septembre 2022

COLLECTION DIGITAL NEW DEAL "NUMÉRIQUE DE CONFIANCE"



SOIT NOUS PARTAGEONS
LES DONNÉES ENTRE NOUS
(EN SE DOTANT D'UNE
INFRASTRUCTURE),
SOIT ELLES RESTENT
CENTRALISÉES PAR
LES BIGTECH.

SOMMAIRE

INTRODUCTION	3
I. LA DATA STRATEGY, INITIATIVE GERMANO-EUROPÉENNE	
1.1. GÉNÈSE D'UN PROJET INITIÉ OUTRE-RHIN.....	6
1.2. EUROPÉANISATION DE L'AMBITION VIA GAIA-X.....	7
II. LA CONFIANCE, PILIER DE NOTRE AUTONOMIE STRATÉGIQUE	
2.1. LES COMPOSANTES ESSENTIELLES DE L'ÉCOSYSTÈME DE CONFIANCE....	12
2.2. QU'EN EST-IL DU CLOUD ?.....	13
2.3. QU'EN EST-IL DE L'IA ?.....	15
2.4. QU'EN EST-IL DE LA DATA ?	16
2.5. CRÉER DE LA VALEUR PAR LE PARTAGE DE CONFIANCE	18
III. INVESTIR LE CHAMP STRATÉGIQUE DES DATA SPACES	
3.1. CONTEXTE EUROPÉEN	20
3.2. CARACTÉRISTIQUES DES DATA SPACE.....	21
3.3. BÉNÉFICES ET CAS D'USAGE	22
3.4. LES DIFFÉRENTS TYPES D'ACTEURS DES DATA SPACES	24
3.5. UN BESOIN FORT DE GOUVERNANCE	28
IV. DÉVELOPPER UNE INFRASTRUCTURE DU PARTAGE DE DONNÉES	
4.1. DÉCENTRALISER L'INFRASTRUCTURE PAR UNE SURCOUCHE D'INTÉROPÉRABILITÉ.....	34
4.2. S'IMPLIQUER PLEINEMENT DANS L'EFFORT DE STANDARDISATION.....	37
4.3. CONSTRUIRE LES COMMUNS NUMÉRIQUES DU PARTAGE DE DONNÉES.....	38
4.4. DÉVELOPPER L'ÉCOSYSTÈME DES ACTEURS DE L'INFRASTRUCTURE.....	40
V. PROPOSITIONS POUR UN PROGRAMME NATIONAL	44
CONCLUSION	46
CONTRIBUTEURS / REMERCIEMENTS	47

INTRODUCTION

'Union européenne s'est toujours définie comme un espace de liberté de circulation : libre circulation des biens, des personnes, des services, des capitaux, et depuis 2018 des données. Cette cinquième liberté nous offre les moyens de renouer avec les origines du numérique, celles du partage, pour nous permettre de nous positionner sur un pied d'égalité avec les géants américains et chinois. Nous n'avons pas véritablement le choix. La donnée est devenue la matière première de l'économie du XXI^{ème} siècle : tout service ou produit, tous domaines confondus, s'il veut être concurrentiel, doit avoir une dimension numérique qui repose sur l'analyse ou l'exploitation de données. L'accès à ces données constitue donc un énorme pouvoir stratégique. Aujourd'hui nous avons d'un côté des géants étrangers qui contrôlent de gigantesques gisements de données, et de l'autre une multitude d'acteurs locaux, publics et privés, qui maîtrisent de petits amas de données, personnelles et non personnelles, souvent inexploitable en l'état. Soit nous partageons les données entre nous, en nous dotant d'une infrastructure logicielle pour les partager et les exploiter, soit elles continueront d'être centralisées par une poignée de géants, captant sans cesse la valeur à notre détriment. C'est ce pari que les réglementations viennent sécuriser (RGPD hier, Data Governance Act et Data Act aujourd'hui), et que les initiatives de partage et d'espaces de données (*data spaces*) viennent déployer (Gaia-X¹).

« Nous définissons aujourd'hui une approche véritablement européenne du partage des données. L'Europe a besoin d'un marché unique des données ouvert mais souverain. Notre réglementation, associée aux investissements adéquats et à des infrastructures clés, aidera l'Europe à se hisser au premier rang mondial dans le domaine des données »

Thierry Breton, dans le contexte du
Data Governance Act, novembre 2020

C'est un enjeu historique, pour l'Europe. Nous avons peu de temps pour gagner la guerre cruciale de la standardisation face aux géants *Big Tech* américains et chinois, en mettant en place pour les filières, dans le cadre Gaia-X en particulier, des standards communs (modèles de données, modèles légaux et business, interopérabilité, identité,...) afin de ne pas dépendre de solutions technologiques extérieures, et de pouvoir aborder offensivement la question de l'innovation.

¹ <https://www.gaia-x.eu/>

Car derrière la question d'une "infrastructure logicielle de partage des données souveraine" se joue la capacité à faire émerger de nouveaux business models innovants basés sur les opportunités que font naître ce changement de paradigme, en résonance notamment avec les potentialités du Web3². **C'est un enjeu majeur pour l'Union européenne, puisqu'il s'agit ici de la construction d'un internet décentralisé, conforme aux valeurs européennes, alternatif aux modèles des plateformes des géants du Web. Ce combat doit absolument être remporté par les européens.** Et si possible, par ceux qui défendent la même définition de l'autonomie stratégique que nous, d'autant que les données sont hébergées dans les systèmes d'information d'entreprises et d'administrations européennes. Cette stratégie doit en effet viser une indépendance par principe vis-à-vis des acteurs extra-européens, et un cadre de confiance par défaut, dans le cas où nous choisissons de nous appuyer sur leurs technologies. C'est pour cela qu'il faut que la France se donne les moyens de rattraper son retard sur l'Allemagne, afin d'une part de rééquilibrer et renforcer le couple franco-allemand, et d'autre part pour promouvoir sa vision politique, singulière, dans le cadre européen. N'ayons aucun doute sur le sujet. Si nous, européens, ne parvenons pas à garantir notre capacité collective à valoriser les données de nos industries et services, d'autres s'en chargeront, de manière très qualitative peut-être, mais au profit exclusif de leurs modèles d'affaire, et au détriment de l'économie européenne, de sa croissance et de son indépendance.

L'objectif de ce *position paper* est d'éclairer le débat sur ces questions complexes. En rappelant tout d'abord l'enjeu du partage des données pour la pleine maîtrise de notre destin technologique et économique, et ensuite en expliquant le plus précisément possible le rôle des *data spaces* dans cette stratégie. **Pour la France, l'enjeu est de définir une "stratégie française de la donnée" s'inscrivant dans la Data Strategy européenne.** La data ne doit pas devenir la grande oubliée de notre ambition du numérique de confiance. **Notre conviction est que seul un État stratège peut penser et impulser un ensemble cohérent de programmes "Cloud, IA et Data",** qui permettra d'offrir aux acteurs métier des offres "de bout en bout" couvrant l'ensemble de la chaîne de valeur numérique de confiance.

² Le Web3 est un terme utilisé pour désigner l'idée d'un web décentralisé exploitant la technologie des chaînes de blocs (blockchain) - Wikipedia



LA FRANCE DÉFINIT
UNE SOUVERAINETÉ
NUMÉRIQUE PAR
LES OUTILS, QUAND
L'ALLEMAGNE
PRÉFÈRE PARLER DE
SOUVERAINETÉ DES
DONNÉES.

I. LA DATA STRATEGY, INITIATIVE GERMANO-EUROPÉENNE

1.1 GENÈSE D'UN PROJET INITIÉ OUTRE-RHIN

Il était une fois Gaia-X... Prenons le temps de raconter ici la genèse de cette initiative afin de mieux comprendre notre situation et de pouvoir ainsi dessiner le rôle que la France peut, et doit, jouer.

Lancée en février 2020, la stratégie européenne pour les données³ se fonde sur la notion d'espaces de données (data spaces) : un concept inédit en matière de politique publique. Un data space⁴ rassemble des acteurs publics et privés désireux de partager leurs données (personnelles et non personnelles) pour créer de la valeur en commun, au moyen d'une infrastructure décentralisée et d'une gouvernance commune. Pour nombre d'acteurs français, c'est une nouveauté qu'ils découvrent avec Gaia-X, fin 2019. Pour les Allemands, c'est une histoire plus ancienne qui démarre en 2015.

La notion de *data space* apparaît dans les années 2000 dans la littérature de recherche académique⁵. *"Les systèmes d'espaces de données offrent des services sur les données sans nécessiter d'intégration sémantique préalable. Contrairement aux systèmes d'intégration de l'information existants, les systèmes data spaces offrent des réponses au mieux avant même que les mappages sémantiques ne soient fournis au système. Les espaces de données offrent une approche de la gestion des données basée sur le principe du "pay-as-you-go". Les utilisateurs (ou les administrateurs) du système décident où et quand il vaut la peine d'investir plus d'efforts dans l'identification des relations sémantiques."*⁶

Concept volontiers "fourre-tout", le *data space* se prête bien aux considérations stratégiques qui cherchent à identifier de grands ensembles de données sans trop les caractériser. Les Allemands vont s'en emparer pour développer leur projet "Industrie 4.0"⁷ destiné à la transformation numérique de leur puissante industrie. C'est le Fraunhofer-Gesellschaft⁸, regroupement d'instituts allemands de recherche appliquée de référence internationale, qui est chargé de lancer l'initiative en octobre 2015. Quelques semaines plus tôt, le gouvernement fédéral avait annoncé son soutien, et des subventions, au profit des travaux du Fraunhofer sur le sujet. À cette même occasion, la création de l'Industrial Data Space Association (IDSA)¹⁰ était également annoncée. Les travaux ne sont pas que théoriques, par vocation le Fraunhofer travaille en lien étroit avec les entreprises, et plus de 60 cas d'usage du *data space* industrie ont déjà été identifiés dans une vingtaine de sociétés.

Boris Otto, l'un de ses chercheurs, publie dès fin 2015 un premier document consacré au *data space* destiné à l'industrie¹¹ (puis en 2016 un livre blanc¹² décrivant l'architecture de référence du *data space* de l'industrie allemande). En janvier 2017, il devient Directeur du Fraunhofer ISST, l'entité de l'institut en charge de l'ingénierie logiciel et des systèmes,

³ https://ec.europa.eu/commission/presscorner/detail/en/fs_20_283

⁴ Les "data spaces" ne doivent pas être confondus avec les "Data hub" pratiqués par les grandes entreprises. Ces derniers n'ont pas de gouvernance ouverte à la concurrence, et n'ont pas vocation à être connectés avec d'autres initiatives dans un cadre politique européen

⁵ <https://en.wikipedia.org/wiki/Dataspaces>.

⁶ *A first tutorial on dataspace* par Michael Franklin, Alon Halevy, David Maier, Aout 2008 ; Voir également, des mêmes auteurs : *From databases to dataspace: a new abstraction for information management*, Décembre 2005

⁷ https://fr.wikipedia.org/wiki/Industrie_4.0

⁸ <https://www.fraunhofer.de/en.html>

⁹ <https://www.plattform-i40.de/IP/Redaktion/DE/Anwendungsbeispiele/277-industrial-data-space-fraunhofer/industrial-data-space.html>

¹⁰ <https://internationaldataspace.org/>

¹¹ *Overview of the industrial data space*, Oct 2015, Fraunhofer, Boris Otto.

¹² *The industrial data space : Architecture blueprint for digital sovereignty*, 2016

qui sera le chef de file de l'initiative, qui rassemblera au total 12 instituts du Fraunhofer.

Toujours très soutenu par le gouvernement fédéral, le Fraunhofer va faire évoluer sa proposition. Le concept de *data space* destiné à l'industrie reste la référence, mais le cadre va être étendu à bien d'autres domaines d'activité. En 2016, *Industrial Data Space Association* (IDSA) va devenir *International Data Space Association*, ce qui marque la volonté d'étendre le concept et également de l'internationaliser. Le Fraunhofer le présente à certains pays européens ainsi qu'à la Commission durant les années 2016 et 2017¹³. L'institut allemand n'a cependant pas limité à l'Europe la promotion de l'IDSA puisqu'il est présenté à l'*International Manufacturing Technology Show* (IMTS) à Chicago en septembre 2018.

Progressivement, la Commission européenne va se rallier au concept de *data space* pour en faire le cœur de sa stratégie. Le sujet ne sera connu par la Commission qu'à partir de 2016. Les travaux allemands du Fraunhofer sont cités dans un document de travail de la DG Connect¹⁴ dès janvier 2017¹⁵, mais c'est la communication de la Commission d'avril 2018 "Vers un espace européen commun de la donnée"¹⁶ qui entérine l'adoption du concept de *data space* avant que la stratégie pour les données ne l'officialise en 2020. Entre-temps, la Commission a réuni une dizaine d'ateliers de travail sectoriels pour préfigurer les *data spaces* correspondants. **La greffe allemande a pris, les *data spaces* sont devenus la grande affaire de la stratégie européenne pour les données.**

Le cadre global est donc fixé. Il s'agit maintenant d'en assurer la déclinaison opérationnelle et industrielle en définissant les architectures et les règles nécessaires, en particulier pour la gouvernance des données, et en promouvant leur circulation puisque c'est la raison d'être des *data spaces*. C'est là qu'intervient Gaia-X. Il revient à l'association AISBL¹⁷, sa maison mère, de prévoir les spécifications techniques et la gouvernance. Les hubs nationaux de Gaia-X doivent pour leur part rassembler aussi largement que possible les producteurs et utilisateurs de données afin qu'ils s'entendent pour déployer, dans chaque secteur, des outils et des plateformes de partage de données, améliorer l'accessibilité, la qualité et de l'interopérabilité des données. Les hubs nationaux doivent appuyer l'AISBL Gaia-X pour soutenir l'émergence de *data spaces* sectoriels, mais doivent dans le même temps veiller à ne pas trop promouvoir une vision nationale des *data spaces*, ce qui serait par exemple un contresens dans de multiples secteurs tels que l'énergie, l'industrie ou la finance.

1.2. EUROPÉANISATION DE L'AMBITION VIA GAIA-X

Afin de lancer le mouvement, les Allemands sollicitent les Français pour monter une équipe de "membres fondateurs" de Gaia-X rassemblant à parité de grandes entreprises des deux pays. Le projet Gaia-X sera initialement annoncé au Digital Gipfel de Dortmund en octobre 2019 en publiant, en français, le document promotionnel¹⁸. La première tâche de l'équipe franco-allemande est de convenir d'un document de position commune, un "*position paper*"¹⁹, qui confirme le partenariat annoncé quelques mois plus tôt par les deux ministres Peter Altmaier et Bruno Le Maire. Ce tandem franco-allemand constitue certainement le signe le plus évident d'un engagement européen, et la meilleure garantie de son succès. Tandem, qui en suit un autre, celui d'Hubert Tardieu et Boris Otto, précurseurs de cette collaboration

¹³ Présentation de IDSA par Fraunhofer aux représentants de la Tchéquie

¹⁴ La direction générale des réseaux de communication, du contenu et des technologies ou Connect est le service de la Commission européenne chargé de la politique de l'Union européenne concernant le marché unique numérique, la sécurité des réseaux, la science et l'innovation numérique.

¹⁵ *Staff Working Document on the free flow of data and emerging issues of the European data economy*, Janvier 2017

¹⁶ Communication de la Commission : *Vers un espace européen commun de la donnée*, Avril 2018

¹⁷ forme juridique d'association internationale à but non lucratif en Belgique

¹⁸ Doc du BMWi : *Le projet Gaia-X - Une infrastructure de données en forme de réseau, berceau d'un écosystème européen vital*, Octobre 2019 doc en français.

¹⁹ *Franco-German Position paper on Gaia-X*, 18 février 2020

franco-allemande à travers le dialogue qu'ils ont initié ensemble, aboutissant à une vision commune préfiguratrice de Gaia-X. Une telle collaboration a pu voir le jour grâce à une culture industrielle franco-allemande initiée notamment par des programmes communs comme celui entre Atos et Siemens sur l'industrie 4.0. Un véritable laboratoire, doté de 320 millions d'euros, qui avait permis de voir se confronter **des visions opposées de l'autonomie stratégique : la France privilégiait la définition d'une souveraineté numérique par les outils**, maîtrisée via des acteurs à minima européens, **quand l'Allemagne préférait parler de souveraineté des données**, sans se préoccuper de la provenance des acteurs en mesure de la garantir...

C'est au travers de ces expériences que Gaia-X a pu être conçue, en se basant sur le consensus issu de ces travaux, et en ambitionnant d'élargir le cadre opérationnel au niveau européen pour tous les secteurs.

En France, depuis la fin de l'année 2019, le Cigref²⁰ s'est engagé dans la mobilisation de ses adhérents afin de participer au Conseil d'administration constitutif de l' AISBL Gaia-X. Des grands groupes comme EDF et Safran ont porté au sein de ce Conseil d'administration les besoins des grands utilisateurs français de services cloud. En novembre 2020, le Secrétaire d'État chargé de la Transition numérique et des Communications électroniques, Cédric O, confiait au Cigref, par lettre de mission, le soin de structurer et d'animer le Hub France de Gaia-X, en partenariat avec l'Académie des technologies, le pôle de compétitivité Systematic Paris-Région, et avec le soutien de la DGE²¹. En 2021, le comité de pilotage du Hub France de Gaia-X s'est élargi à l'INRIA²², Numeum²³ et l'IMT²⁴. Le Hub France de Gaia-X a ainsi organisé, structuré et animé 12 groupes de travail sectoriels dont la plupart sont désormais inscrits dans la dynamique européenne portée par l' AISBL Gaia-X.

Côté allemand, le projet phare illustrant la mise en œuvre industrielle des *data spaces* est Catena-X²⁵, qui rassemble au départ les géants de l'industrie automobile allemande comme BMW, Mercedes, Volkswagen, ainsi que de multiples sous-traitants.

En six ans, la vision initiale d'un projet destiné à l'origine à garantir la suprématie de l'industrie allemande au temps du numérique 4.0 a évolué pour devenir une politique publique nationale, puis européenne, tout en aboutissant à un projet d'ambition mondiale (le Japon et la Corée ont rejoint Gaia-X en 2022). Les plus grandes entreprises du numérique dans le monde sont aujourd'hui membres de Gaia-X. L'histoire de cette initiative illustre remarquablement la réussite de la méthode allemande, celle de la construction concertée, structurée, maîtrisée des *data spaces*, sous l'égide de l'association du gouvernement fédéral avec le Fraunhofer qui, depuis 2015, n'a cessé de rassembler des entreprises et des associations autour d'elle, qu'elles soient allemandes, européennes ou de rang mondial.

Ce succès européen (+20 pays) n'est cependant ni partagé, ni même vraiment reconnu en France, où Gaia-X suscite davantage de questions que d'enthousiasme. Sur la même période, une stratégie française sur les données a émergé avec succès, mais avec une orientation différente, en l'occurrence très nationale, et surtout davantage tournée vers l'ouverture (Open Data) que le partage de données (Shared Data). Les deux approches pourraient parfaitement être complémentaires.

²⁰ <https://www.cigref.fr/>

²¹ Direction générale des entreprises

²² Institut national de recherche en informatique

²³ Fusion de Syntec Numérique et Tech in France

²⁴ Institut Mines-Télécom

²⁵ <https://catena-x.net/en>

OPEN DATA**LA POLITIQUE FRANÇAISE DE L'EXPLOITATION ET L'OUVERTURE DES DONNÉES PUBLIQUES**

Depuis le début de la décennie 2000, le gouvernement français n'a eu de cesse de valoriser et de mettre à disposition du plus grand nombre des données publiques en s'appuyant sur le remarquable patrimoine de données que les services publics ne cessent d'alimenter. Cette ambition a constitué dès 2014 une politique publique confiée à l'Administrateur général des données de l'État²⁶. Elle a été confirmée et renforcée en 2016 par la Loi pour une République numérique²⁷.

*Plus récemment, le rapport sur "la politique publique de la donnée, des algorithmes et des codes sources" confié au député Éric Bothorel²⁸, remis en fin 2020, a relancé la même politique enrichie d'un chapitre nouveau sur l'ouverture du code et des algorithmes. Nous pouvons être fiers de cette stratégie nationale, largement inspirée du concept d'État-Plateforme²⁹, mais il faut savoir qu'elle trouve assez peu d'écho dans les autres pays européens comme l'Allemagne notamment, puisqu'ils ne disposent pas des mêmes patrimoines de données publiques. **Très tournée vers l'État et ses relations avec les citoyens/usagers, l'Open Data ne peut pas constituer à lui seul un levier de politique économique de la donnée à l'intention du secteur privé.***

En effet, une stratégie de la donnée ne peut se concentrer uniquement sur les données qu'on peut ouvrir à tous (l'Open Data ne couvre pas les données personnelles, ou plus sensibles d'un point de vue commercial, alors que ces données intéressent tout particulièrement les filières économiques).

Si beaucoup d'entreprises ont rallié le Hub France de Gaia-X dans ses premiers temps, trop peu s'engagent véritablement et la structure, qui ne dispose d'aucun moyen ou ressource en propre, peine à passer d'une fonction de préfiguration des data spaces sectoriels à une fonction d'accompagnement de leur opérationnalisation. La plupart des groupes de travail continuent cependant de travailler et de développer leur activité au niveau européen, notamment grâce à l'engagement, souvent remarquable, de leurs préfigurateurs. Le hub peine cependant, par manque de moyens, à atteindre la pleine puissance escomptée. Par ailleurs, ce manque de ressources ne permet pas au Hub France de Gaia-X de s'investir dans les groupes de travail de l' AISBL Gaia-X, et d'animer et coordonner ses travaux au niveau national.

Pourtant, ailleurs en Europe, les ralliements à Gaia-X se multiplient. Près d'une vingtaine de pays ont créé leur propre hub national. L'association Gaia-X AISBL installée à Bruxelles voit le nombre de ses adhérents augmenter régulièrement et réussit sa montée en puissance depuis sa création au début 2021. Son rapprochement avec la Commission est acquis, ce qui permet de valider le partenariat indispensable entre le niveau politique, la DG Connect³⁰, et celui des opérations, assuré par Gaia-X.

En ce printemps 2022, l'actualité du cloud et de la data en France est dominée par l'AWS³¹ Summit qui a réuni près de 7000 personnes le 12 avril. Cet engouement pour le plus gros

²⁶ Le décret du 16 septembre 2014 institue, auprès du Premier ministre, un administrateur général des données (AGD), rattaché au secrétaire général pour la modernisation de l'action publique (SGMAP).

²⁷ Loi dite "Lemaire" du nom de la secrétaire d'État chargée du numérique et de l'innovation Axelle Lemaire, promulguée le 7 octobre 2016

²⁸ <https://www.gouvernement.fr/actualite/remise-du-rapport-sur-la-politique-publique-de-la-donnee-des-algorithmes-et-des-codes-sources>

²⁹ La notion d'État plateforme apparue dans l'essai de Tim O'Reilly, *Government as a Platform*, montrant l'intérêt pour l'État à l'ère numérique d'adopter des stratégies de plateformes. Concept popularisé à partir de 2012 par l'ouvrage de Nicolas Colin et Henri Verdier, *L'Age de la multitude, entreprendre et gouverner après la révolution numérique*.

³⁰ *La direction générale des réseaux de communication, du contenu et des technologies est le service de la Commission responsable de la politique de l'UE concernant le marché unique numérique, la sécurité sur internet, ainsi que la science et l'innovation numériques.*

³¹ Amazon Web Service

des hyperscalers³² américains n'est pas nouveau. Entre 2020 et 2021, de multiples grandes entreprises françaises ont signé d'importants contrats ou des partenariats avec AWS, Microsoft Azur ou Google Cloud³³. **Dans les conditions juridiques actuelles, aucune d'entre elles ne dispose de garantie d'une quelconque forme de souveraineté sur ses données,** pas plus que de possibilité d'interopérabilité, de réversibilité ou de portabilité. **C'est pourtant ce que propose Gaia-X³⁴** afin de faciliter et sécuriser les échanges de données pour permettre de créer de la valeur.

Les acteurs français vont-ils durablement rester à l'écart de ce puissant mouvement européen de développement de l'économie de la donnée ?

« Il est important que les entreprises françaises et notamment les entreprises propriétaires et utilisatrices des données se mobilisent, avec le soutien de l'Etat, pour créer entre elles des écosystèmes créateurs de valeur, comme ont su le faire en Allemagne nombre de filières dont notamment l'automobile. Nous appelons de nos vœux ces entreprises à rejoindre les groupes de travail Gaia-X pour peser et amener leur expertise utile et nécessaire face à des entreprises extra-européennes très puissantes »

Martine Gouriet, EDF,
Data Space Energie

³² *hyperscaler* : l'hyperscale est le procédé qui consiste à mutualiser les ressources serveurs (cloud computing). Un hyperscaler est une entité qui propose ce service comme Scaleway, Switch, Alibaba, IBM, QTS, Digital Realty Trust, Equinix, Oracle, Facebook, Amazon Web Services, SAP, Microsoft ou Google, etc.

³³ En 2021, Safran, Stellantis, Total, SNCF, ENGIE a engagé un partenariat avec AWS, Société Générale et ArcelorMittal avec Microsoft Azure, LVMH avec Google Cloud. En 2020, Renault, Carrefour, Sanofi et Orange ont signé avec Google Cloud.

³⁴ En 2021, sous l'impulsion des parties prenantes françaises, l' AISBL Gaia-X s'est par exemple engagée dans une démarche de traduction de ses policy rules dans des labels applicables aux services cloud.



POINT DE LIBERTÉ
DE CIRCULATION
SANS CADRE
DE CONFIANCE.

II. LA CONFIANCE, PILIER DE NOTRE AUTONOMIE STRATÉGIQUE

2.1 LES COMPOSANTES ESSENTIELLES DE L'ÉCOSYSTÈME DE CONFIANCE

Point de liberté de circulation, sans cadre de confiance.

L'Union européenne doit poursuivre son effort stratégique de protection des données. Après les données personnelles avec le RGPD, notre continent doit maintenant étendre ces garanties à la protection des données non personnelles sensibles de ses industries et services. Et il y a urgence en la matière car jamais dans l'histoire, les entreprises n'avaient généré et manipulé autant de données, situant les enjeux d'intelligence économique à un niveau encore jamais atteint.

Rappelons ici que l'arrêt de la CJUE³⁵ du 16 juillet 2020 relatif à l'invalidation du Privacy Shield met en exergue la section 702 du Foreign Intelligence & Surveillance Act³⁶ qui autorise les agences de renseignement américaines, notamment la NSA³⁷, de pratiquer la collecte massive, a priori et sans mandat des données de toutes les "non-US persons", qu'elles soient physiques ou morales. Or, une partie substantielle de l'activité des agences de renseignement consiste à effectuer du renseignement d'intérêt économique, notamment au profit des entreprises américaines, dans le cadre d'un plan de renseignement piloté et coordonné au niveau fédéral. Au titre de cet article 702 du FISA, et probablement dans le cadre d'autres législations analogues comme le Patriot Act, **les entreprises américaines de technologie, notamment les hyperscalers, sont tenues de mettre, à leurs frais et de manière secrète, à la disposition des agences de renseignement, les moyens permettant d'accéder aux données des cibles non-US de ces agences.** Le cloud, dans sa dynamique exponentielle sur le marché européen, et la captation hégémonique de celui-ci par les hyperscalers, offre à des agences de renseignement non européennes un moyen, sans équivalent dans le passé, d'accéder massivement aux données sensibles de l'économie de notre continent. Face à cette menace, l'Union européenne doit se doter de mécanismes permettant de garantir la libre circulation des données sensibles des organisations publiques et privées européennes, dans des conditions de confiance vérifiables et opposables.

Point d'autonomie stratégique, sans écosystème de confiance

En matière de numérique, comme de militaire ou d'énergie, la souveraineté est une ambition politique, l'autonomie stratégique un objectif, et le niveau de dépendance une mesure. La souveraineté numérique d'un État, peut se définir en fonction de son niveau d'autonomie, c'est-à-dire sur son aptitude à choisir, et non pas subir, le degré et la nature de ses dépendances : *"L'autonomie stratégique est une capacité à générer et défendre un écosystème de confiance qui organise nos interdépendances technologiques"*³⁸.

Pas de souveraineté numérique sans écosystèmes de confiance qui garantissent l'autonomie stratégique de nos infrastructures de cloud, des données qui s'y trouvent (ou y sont partagées), et des technologies qui les exploitent (à base d'IA notamment). Pour garantir cette autonomie technologique et juridique, et donc demain politique et économique,

³⁵ Cour de justice de l'Union européenne

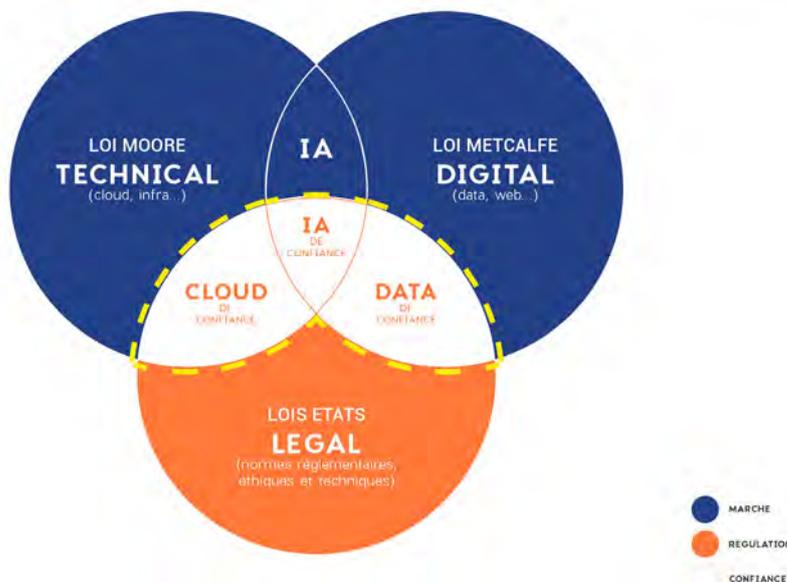
³⁶ <https://www.nsa.gov/Signals-Intelligence/FISA/>

³⁷ National Security Agency (USA)

³⁸ IA de confiance : une opportunité stratégique pour la souveraineté industrielle et numérique, Julien Chiarini et Arno Pons, Digital New Deal, juin 2022.

il faut s'assurer que nous maîtrisons toutes nos dépendances sur l'infrastructure matérielle comme logicielle, qui sous-tend le triptyque Cloud-Data-IA (cf. schéma ci-dessous, matrice de notre collection de publications "Numérique de confiance" dont fait partie cette note)³⁹.

ÉCOSYSTÈME DE CONFIANCE (CLOUD + DATA + IA) = AUTONOMIE STRATÉGIQUE



Cet écosystème de confiance voit sa traduction opérationnelle dans les *data spaces*, puisqu'ils couvrent le périmètre de ses trois composantes Cloud-Data-IA (et non pas seulement data comme son nom pourrait à tort le laisser penser...). **Mener la bataille des *data spaces* de confiance, c'est conquérir notre autonomie stratégique.**

Pour mettre en place une telle infrastructure, nous avons besoin d'une prise de conscience généralisée, d'une stratégie globale d'État et de financements cohérents. Pour chacune des trois composantes - Cloud, Data, et IA - un programme dédié à la coopération est souhaitable.

2.2 QU'EN EST-IL DU CLOUD ?

Le cloud est souvent considéré comme la mère de toutes les batailles. Plus qu'une simple commodité, il est devenu pour les acteurs publics et privés, à tort ou à raison, la clé d'entrée de leurs architectures stratégiques digitales. Les *hyperscalers* ont profité de leur position dominante pour accentuer ce phénomène en faisant du cloud l'infrastructure matricielle du numérique, enchevêtrant les couches, et agrégeant volontairement les services, au point de ne plus rendre visible sa composition technique. Les *hyperscalers* ont réussi à imposer dans l'imaginaire collectif que le *Big Data* ne pouvait être géré que par leurs services cloud, et que tout autre choix serait technologiquement et économiquement non viable.

C'est ainsi que les géants américains ont siloté l'ensemble des chaînes de valeur via des offres "packagées" Cloud-Data-IA, enfermant ainsi leurs clients dans des systèmes propriétaires dont on ne peut facilement, ni sortir (très complexe et coûteux), ni même mixer avec des services concurrents... C'est ce que nous dénonçons en 2021 dans notre livre blanc "Cloud de confiance" lorsque nous avons identifié des actions de concurrence déloyale via les ventes liées : "En subordonnant l'achat d'un service à l'achat

³⁹ Digital New Deal publie une collection de quatre notes sur la définition du "numérique de confiance" comme socle de notre autonomie stratégique: *Cloud de confiance* par Laurence Houdeville et Arno Pons en 2021, puis en 2022 *Cybersécurité, vigile de notre autonomie stratégique* par Arnaud Martin et Didier Gras, *Data de confiance* par Olivier Dion et Arno Pons, et *IA de confiance* par Julien Chiaroni et Arno Pons.

d'autres, certains fournisseurs de Cloud entravent la vente de services connexes par des entreprises concurrentes. Ces pratiques rendent opaque l'offre à partir du moment où chaque service vendu ne bénéficie pas d'un prix associé. Des obligations en matière de vente subordonnée ou dite "conjointe" pourraient être édictées à commencer par une lecture individualisée des services et de leurs prix, compris dans un package. Ceci permettrait aux entreprises de choisir ou non l'ensemble de l'offre en bénéficiant de tarifs modulaires.⁴⁰

CAPTATION ET SILOTAGE DE LA VALEUR



Pour sortir de ce "Syndrome de Stockholm numérique"⁴¹ et déverrouiller ce marché hautement stratégique, il faut donc d'une part imposer une transparence commerciale et juridique, et d'autre part oser réintermédier le marché. A cet agrégat monolithique, **nous devons en effet réussir à imposer une couche intermédiaire facilitant la transversalité.**

Il y a urgence à agir car les attermoissements de l'Europe renforcent inéluctablement les positions acquises par les offres dominantes de cloud public portées par les *hyperscalers*. A court terme, les entreprises et les administrations publiques n'auront d'autre choix, compte tenu de leurs échéances contractuelles, des besoins de compétitivité et de service, et des choix technologiques en faveur du cloud des principaux grands éditeurs, que d'arbitrer à nouveau et durablement au profit des *hyperscalers*. Elles réduiront ainsi à une très faible proportion les cas d'usage dont pourrait bénéficier une offre nationale ou européenne de cloud de confiance et de maîtrise de la valorisation des données. Dans ces conditions, l'Union européenne a besoin de manière urgente d'un cadre législatif et opérationnel pour maîtriser et garantir la confiance de cette cinquième liberté de circulation des données sur notre continent. Le Data Act doit être conçu pour apporter ce cadre législatif et Gaia-X en développer le cadre opérationnel et industriel.

⁴⁰ Cloud de confiance, en enjeu d'autonomie stratégique pour l'Europe, Laurence Houdeville et Arno Pons, Digital New Deal, 2021

⁴¹ Syndrome de Stockholm numérique, Jean-Romain Lhomme, Digital New Deal, 2018

STRATÉGIE NATIONALE POUR LE CLOUD

Afin de soutenir l'innovation des acteurs nationaux et européens, la France s'est dotée à la fin de l'année 2021 d'une **stratégie industrielle nationale pour le cloud**⁴².

Ce troisième volet de la stratégie nationale d'accélération cloud, après la définition du "cloud de confiance" et la doctrine "cloud au centre" pour les services de l'État, doit permettre de mobiliser **1,8 milliards d'euros de soutien à l'écosystème cloud français sur quatre ans**, soit 667 M€ de financements publics, issus du Programme d'investissements d'avenir (PIA), ajoutés à 680 M€ de financements privés et 444 M€ de financements européens. Parmi ses priorités :

- mise en avant des offres françaises innovantes
- aider les acteurs français à passer à l'échelle sur les principales technologies critiques telles que le *Big Data* ou le travail collaboratif
- développer les technologies de rupture à l'horizon 2025 (edge computing,...)

Cette politique de soutien à l'offre s'appuie :

- sur les principaux acteurs nationaux (OVHcloud, Outscale, Oodrive, Scaleway,...)
- au niveau européen sur Gaia-X afin de financer la conception des services fédérés GXFS au cœur de l'émergence de *data spaces* de confiance, et la mise en place d'un schéma européen de certification de cybersécurité pour les services cloud

2.3. QU'EN EST-IL DE L'IA ?

Par leur domination sur le marché du cloud, au travers duquel ils proposent déjà des outils de développement d'Intelligence Artificielle (MLOps⁴³, etc.), et sur les applications utilisatrices d'Intelligence Artificielle, **les hyperscalers sont parfaitement placés pour prendre et verrouiller le marché des outils et méthodes de développement des technologies à base d'IA**. Conséquence pour l'écosystème de confiance que l'Europe souhaite mettre en place, **les hyperscalers risquent de "privatiser" la confiance dans l'IA**, pourtant un élément majeur de la stratégie numérique européenne⁴⁴ et de sa future régulation (AI Act).

Pour répondre à cet enjeu, la France dispose déjà d'un programme de coopération national avec le Grand Défi IA (porté par le Secrétariat Général Pour l'Investissement - SGPI) et sa déclinaison opérationnelle "Confiance.ai" qui adresse concrètement le besoin de mutualisation d'investissement entre les filières pour la confiance dans l'IA.

⁴² [Présentation à la presse le 2 novembre 2021](#)

⁴³ MLOps ou ML Ops est un ensemble de pratiques qui vise à déployer et maintenir des modèles de machine learning en production de manière fiable et efficace

⁴⁴ *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

EXEMPLE DU GRAND DÉFI IA ET DU PROGRAMME CONFIANCE .AI

L'investissement de la France pour l'IA de confiance est porté par le SGPI (Secrétariat Général Pour l'Investissement) aux travers du Grand Défi "Sécurisation, fiabilisation et certification des systèmes à base d'intelligence artificielle", qui s'articule autour de 3 piliers :

- **Pilier #1 Infrastructure** (le programme Confiance.ai) qui permettra de développer un environnement de conception de systèmes à base d'IA de confiance : algorithmes IA, design des systèmes, déploiement et maintenance des systèmes critiques basés sur l'IA, etc.
- **Pilier #2 Conformité** qui permettra d'assurer la bonne conduite opérationnelle de l'exploitation des systèmes reposant sur l'IA de confiance, puis de définir le rôle et les compétences des certificateurs/ tiers de confiance.
- **Pilier #3 Normes** qui permettra d'établir, en concertation avec les différents acteurs de l'industrie, les normes, standards, environnement réglementaire et certifications.



Cet investissement permet à la France de développer une infrastructure de l'IA de confiance s'appuyant sur les valeurs et la réglementation européenne. On constate que les enjeux de normalisation sont une partie intégrante du programme, qui, par ailleurs, fédère et coordonne toutes les parties prenantes (acteurs publics, industriels, grands groupes, startups et académiques). Si le programme IA de confiance s'appuie pour le moment principalement sur des acteurs des secteurs des systèmes critiques et à haut risque, **les développements IA de confiance pourraient toutefois être adaptés aux data spaces** dans tous les secteurs (éducation, agriculture, mobilité, tourisme,...).

Le *position paper franco-allemand "speeding up industrial AI and trustworthiness"*⁴⁵, fruit du partenariat entre le Secrétariat Général pour l'Investissement (SGPI) et la *Big Data Value Association (BDVA)* en collaboration avec une cinquantaine d'experts académiques et industriels, illustre également le **lien étroit entre confiance et industrialisation de l'IA**. Il souligne le **besoin industriel de solutions logicielles** pour implémenter la confiance dans les processus et applications à base d'IA, **couplé à la disponibilité de normes et de standards, ceci en complémentarité avec les initiatives européennes de type Gaia-X** et d'autres initiatives sur l'IA.

2.4. QU'EN EST-IL DE LA DATA ?

La confiance dans la data est encadrée en Europe par un ensemble de règlements (RGPD, Data Governance Act, Data Act, Digital Services Act, Digital Markets Act) **ayant pour vocation à protéger les individus, comme les organisations**. Pour être réellement efficace, **la régulation ne peut pas suffire, elle doit impérativement être associée à une stratégie industrielle** permettant de déployer ces règlements, ainsi que les normes et standards associés, dans des applications métier. **Si nous n'agissons pas, l'Europe risque de voir les hyperscalers imposer des standards de facto**.

Gaia-X a pour mission de proposer des standards et labels, en cohérence avec les textes réglementaires européens, pour les data spaces. Ces standards sont traduits dans des **Communs Numériques, dont le socle de base est appelé GXFS**⁴⁶ (*pour Gaia-X Federation Services*). Les GXFS sont actuellement en cours de développement à l'échelle européenne, ils sont composés de 4 briques : *Identity & Trust, Data Sovereignty Services, Federated Catalogue et Compliance*.

⁴⁵ "speeding up industrial AI and trustworthiness", Secretariat Général pour l'Investissement, Big Data Value Association et al., 2021

⁴⁶ <https://www.gxfs.eu/gxfs-overview/>

Grâce aux GXFS, chaque *data space* Gaia-X pourra construire son infrastructure de partage de données en se basant sur des éléments open source communs à tous les projets, facilitant ainsi l'interopérabilité à grande échelle. Les GXFS permettront à tous les acteurs du réseau de se décrire de manière similaire (son identité, les données mises à disposition du réseau, son niveau de conformité aux règles/labels Gaia-X, etc.).

GAIA-X FEDERATION SERVICES (GXFS)

UN PREMIER FINANCEMENT POUR LES COMMUNS NUMÉRIQUES DES DATA SPACES

Les GXFS français ont été financés pour en contrepartie des GXFS allemands qui écrivaient jusqu'ici presque seuls les spécifications...En Allemagne le projet GXFS est porté par eco⁴⁷, en France il est porté par l'Institut Mines-Télécom (IMT) dans le cadre de l'AMI Cloud Souverain. **Le projet français correspond à un financement de 12M€ qui associe quelques membres fondateurs de Gaia-X** (ATOS, Dawex, Docaposte, Outscale, OVH). Il s'étend sur 24 mois, avec un démarrage en juillet 2021.

Le périmètre du projet français comporte l'amorçage des développements open source, et l'écriture des spécifications des GXFS. L'IMT via TeraLab⁴⁸ servira de "testbed"⁴⁹ pour valider la conformité des projets par rapport à Gaia-X, en s'appuyant sur ces GXFS.

Les GXFS sont essentiels au bon fonctionnement des *data spaces*, ils sont le liant de l'ensemble de l'écosystème. Cependant le projet français n'en couvre qu'une partie et il faudra ensuite "faire vivre" ces GXFS dans les infrastructures de partage de données des projets métier. Ils permettent simplement à chaque *data space* de partir sur une base commune, et ne représentent au final qu'une partie infime du volume des développements nécessaires à la création complète de l'infrastructure des *data spaces* ; ce qui implique la **nécessité d'encourager le développement d'une forte communauté open source autour du projet, et l'augmentation du nombre d'acteurs "InfraTech"** pour constituer l'écosystème qui sous-tend à la structure de Gaia-X.

La France a donc déjà posé les bases d'un potentiel programme "Data de confiance" en investissant dans le développement des GXFS, socle open source de Gaia-X. C'est donc un passage à l'échelle auquel nous appelons, afin que ces efforts de fédération ne restent pas vains. Il est également primordial de lancer et d'animer une communauté open source pour faire évoluer les GXFS (la France et l'Allemagne ne vont en développer qu'une partie) et proposer un support associé, des services d'intégration et des offres de maintenance. Le financement de cette "boîte à outil pour créer des fédérations" est un atout majeur dans notre capacité opérationnelle "d'empowerment" des *data spaces*. **Mais cet investissement n'atteindra sa pleine capacité que s'il s'inscrit dans un programme de coopération complet en France pour la création d'une infrastructure de partage de données, et l'animation des *data spaces*, qui vont s'appuyer sur ces GXFS.** En effet, il faut également soutenir les initiatives sectorielles (éducation, compétences, tourisme, énergie, mobilité, etc.) qui vont mettre en œuvre de manière concrète ces *data spaces*, et ainsi contribuer aux communs généraux aussi bien qu'au développement de briques sectorielles.

La data est pourtant la véritable clé de voûte de notre écosystème de confiance numérique car c'est principalement par la data que la valeur se crée. **Il n'existe pas de "Grand Défi Data" qui viendrait sécuriser les investissements faits sur le Cloud de confiance, sur l'IA de confiance, et sur les GXFS.** Or, pour peser dans Gaia-X et ne pas subir des choix structurants déjà en cours, nous avons besoin d'une plus grande coordination des moyens. Ce programme "Data

⁴⁷ Avec plus de 1 000 entreprises membres, eco est la plus grande association de l'industrie de l'Internet en Europe. Depuis 1995, eco a joué un rôle déterminant dans l'évolution de l'Internet, en encourageant les nouvelles technologies, en créant des conditions cadres et en représentant les intérêts de ses membres dans la politique et les comités internationaux. Les thèmes clés d'eco sont la fiabilité et le renforcement de l'infrastructure numérique, la sécurité et la confiance dans les technologies de l'information, ainsi que la numérisation orientée vers l'éthique.

⁴⁸ <https://www.teralab-datascience.fr/>

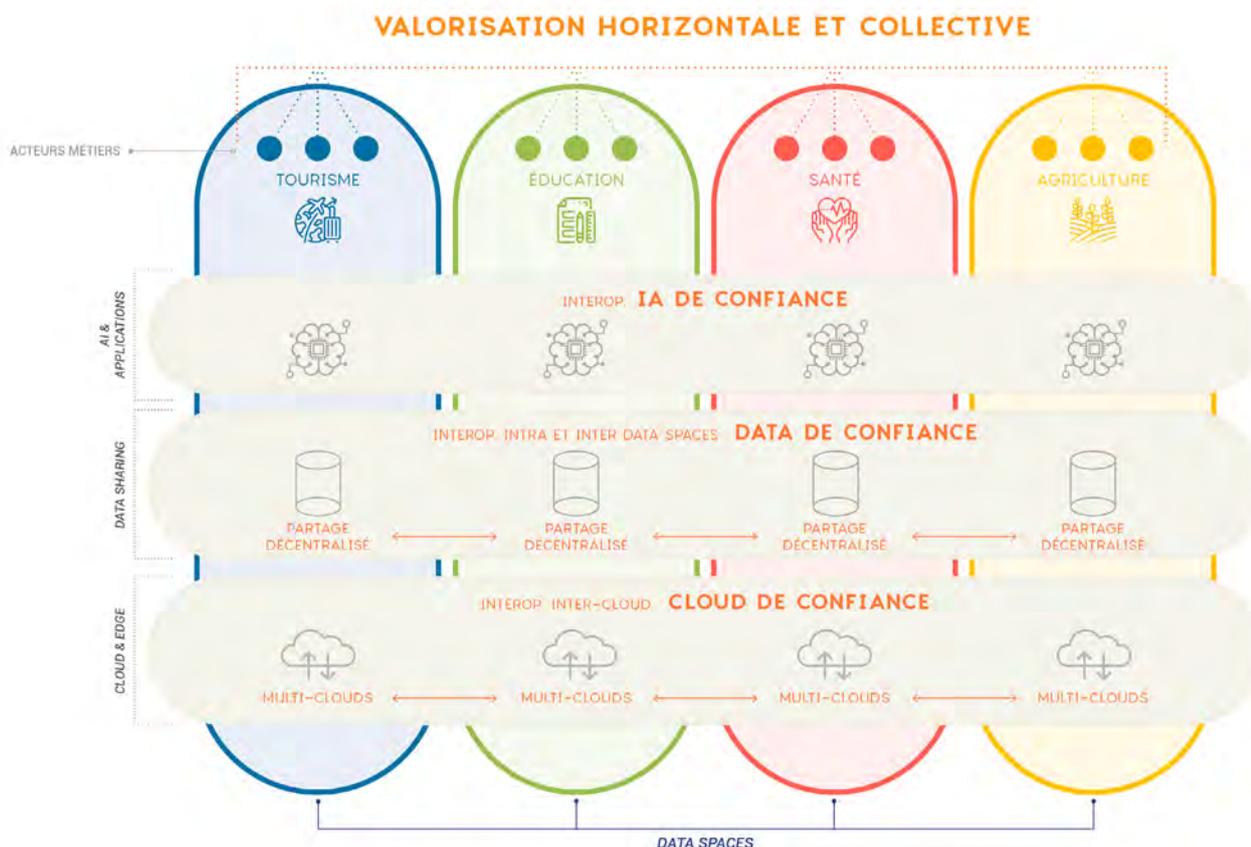
⁴⁹ Plateforme de tests

de Confiance" devrait être piloté en cohérence avec les politiques en cours sur Cloud et l'IA, afin de maintenir une feuille de route commune sur ce triptyque de confiance "Cloud-Data-IA", tout en accordant à la data une place à part. Ceci est d'autant plus stratégique que **la data est paradoxalement le terrain d'innovation le plus propice pour l'Europe. Il nous sera en effet plus facile de régler les problèmes de gouvernance du partage des données, que de combler directement notre retard technologique et commercial sur le cloud et l'IA face aux hyperscalers.**

2.5. CRÉER DE LA VALEUR PAR LE PARTAGE DE CONFIANCE

Les **data spaces** doivent permettre de briser les barrières qui enferment les acteurs économiques dans des silos qui permettent aux *hyperscalers* d'être leurs interlocuteurs privilégiés sur l'ensemble de la chaîne de valeur Cloud-Data-IA, captant ainsi la valeur de manière verticale.

Les couches d'interopérabilité Cloud-Data-IA que l'Europe est en train de créer, doivent permettre une **horizontalisation des échanges, et donc une collectivisation de la valeur.** Songeons ici aux applications concrètes que cela représente : "Prenons l'exemple de la santé. Imaginons que Philips (client d'AWS) veuille travailler avec Sanofi (client de Google Cloud) : est-ce que l'interopérabilité des données sera possible ? Et si oui, est-ce que leurs conditions de partage de données seront réellement respectées ?"⁵⁰



⁵⁰ Hubert Tardieu, Conference on European Digital Sovereignty, Munich, mai 2022



LES COUCHES
D'INTEROPÉRABILITÉ
ET DE MUTUALISATION
CLOUD-DATA-IA
DOIVENT PERMETTRE UNE
HORIZONTALISATION DES
ÉCHANGES, ET DONC
UNE COLLECTIVISATION
DE LA VALEUR.

III. INVESTIR LE CHAMP STRATÉGIQUE DES DATA SPACES

3.1 CONTEXTE EUROPÉEN

La stratégie Européenne des données⁵¹ (*data strategy*) comprend un ensemble de textes réglementaires répondant à la vision énoncée par la Présidente de la Commission Ursula von der Leyen : « les technologies numériques doivent enrichir notre vie à tous et respecter les valeurs européennes ». Dans cet esprit, le Data Governance Act (DGA) et le Data Act (DA) mettent en place le cadre de développement du partage de données (*data-sharing*), ouvrant la voie au développement de multiples *data spaces*.

En 2022, la Commission européenne a lancé un programme d'investissement de presque 10 milliards d'euros sur le sujet des "*Common European data spaces*" au cœur de la *data strategy*, en complément des investissements réalisés par les États membres et le secteur privé. L'ambition de l'Europe est de construire une "*soft infrastructure*" du partage des données basée sur les valeurs et les réglementations européennes : RGPD, Data Governance Act (DGA), Data Act (DA), Digital Services Act (DSA), Digital Markets Act (DMA), AI Act. Cette infrastructure devra favoriser l'émergence d'un modèle alternatif à celui des *hyperscalers* (centré sur les plateformes), en faveur d'écosystèmes de partage de données décentralisés, équitables et centrés sur l'humain. Concernant le partage des données personnelles, l'UE rappelle que le RGPD, avec ses valeurs, ses principes et ses droits, sont les éléments constitutifs de l'infrastructure qui sera mise en place.

L'UE favorise l'émergence de projets de *data spaces*, ainsi que de gouvernances sectorielles qui seront chargées de coordonner des secteurs clés pour faciliter la circulation des données sur la base de normes communes. Dix secteurs prioritaires ont été annoncés à ce jour dont la mobilité, la finance, la santé, les compétences, l'énergie, le *Green Deal* et l'administration.

La taille de marché globale pour le partage de données est estimée à 530 Milliards €⁵² par an par la Commission.

Le cadre de standardisation des *data spaces* est en cours de construction au sein d'organisations telles que aNewGovernance⁵³, BDVA⁵⁴, Fiware⁵⁵, Gaia-X, IDSA, MyData⁵⁶, et autres. Il sera formalisé via le DGA, complété par le DA, dont l'objectif principal est de « favoriser la disponibilité des données en vue de leur utilisation, en augmentant la confiance dans les intermédiaires de données et en renforçant les mécanismes de partage des données dans toute l'UE ».

Ce texte prévoit par exemple la mise en place des :

- **Data Innovation Board** composé de représentants de tous les États membres, de l'EDPB (European Data Protection Board), des *data spaces* pertinents et d'autres représentants des autorités compétentes dans des secteurs spécifiques. Il apportera son expertise lorsque des décisions sur la gouvernance des données devront être prises. Le conseil conseillera la Commission sur la cybersécurité pour les échanges et le stockage de

⁵¹ <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52020DC0066&from=EN>

⁵² Commission européenne (2020): *Support Study to this Impact Assessment*, SMART 2019/0024, Deloitte

⁵³ <https://www.anewgovernance.org/>

⁵⁴ <https://www.bdva.eu/>

⁵⁵ <https://www.fiware.org/>

⁵⁶ <https://www.mydata.org/>

données, et poussera en faveur de la normalisation internationale et de la portabilité des données.

- **Data Space Support Center (DSSC)** qui coordonne toutes les actions pertinentes sur les espaces de données sectoriels et met à disposition des architectures et des exigences en matière d'infrastructure de données pour les *data spaces*, y compris les technologies, processus, normes et outils possibles qui permettront la réutilisation des données entre les filières par le secteur public et les entreprises européennes, notamment les PME.

Enfin, **la Commission européenne investit sur la mise en place d'un *smart middleware*⁵⁷ open source, au-delà des GXFS, correspondant à la couche intermédiaire d'infrastructure commune à tous les *data spaces***. L'Europe affiche donc clairement une volonté de création de Communs Numériques facilitant l'émergence des *data spaces* où l'action coordonnée des États membres est nécessaire.

La France, avec ses capacités technologiques, ses grands groupes et startups innovantes, ainsi que sa volonté politique, est tout à fait en mesure de se positionner comme l'un des leaders de cette démarche en s'insérant pleinement dans le projet européen.

3.2. CARACTÉRISTIQUES DES DATA SPACES

Afin de mieux comprendre les enjeux associés à l'émergence des *data spaces*, en voici les fonctionnalités principales :

- **Une fédération d'infrastructures informatiques sécurisées pour mettre en commun, accéder, traiter, utiliser et partager les données** (personnelles, non-personnelles, industrielles) de manière décentralisée et contrôlée.
- **Un mécanisme de gouvernance des données**, comprenant un ensemble de règles de nature législative, administrative, technique et contractuelle qui déterminent les droits d'accès, de traitement, d'utilisation et de partage des données fiables et transparents.
- **Le contrôle de l'accès aux données par les fournisseurs de données, et les citoyens**, qui décident à quelles fins et dans quelles conditions elles peuvent être utilisées, notamment via des intermédiaires de données, tels que décrits dans le Data Governance Act.
- **La réutilisation de grandes quantités de données**, qui peuvent être réutilisées sous certaines conditions, contre rémunération ou gratuitement, et de manière contrôlée par le fournisseur de données et/ou le citoyen.
- **La participation ouverte d'un grand nombre d'organisations et d'individus.**

Ces fonctionnalités sont mises en œuvre au travers de quatre services de fédération :

- **Le catalogue fédéré** : qui fonctionne comme un registre de toutes les ressources disponibles sur le réseau. Les entrées du catalogue comprennent des descriptions des ressources mêmes (telles que les données et les services), ainsi que les politiques auxquelles les services adhèrent (par exemple le RGPD), et les conditions dans lesquelles les ressources peuvent être utilisées par d'autres participants au réseau.
- **L'échange sécurisé / de confiance de données** : qui permet la définition, l'échange, le traitement et le suivi des politiques d'utilisation des données. Ceux-ci peuvent être

⁵⁷ <https://digital-strategy.ec.europa.eu/fr/news/simpl-cloud-edge-federations-and-data-spaces-made-simple>

compris comme les termes et conditions de l'économie des données. Les exemples sont la limitation du nombre d'actions de lecture sur les données, l'interdiction de distribuer davantage les données, l'obligation d'utiliser les données uniquement dans une certaine zone géographique, etc. Cela concerne également la gestion du consentement des individus sur le partage et traitement de leurs données.

- **La gestion des identités** : qui garantit la confiance entre les fournisseurs et les utilisateurs du réseau (organisations et individus).
- **Les services de conformité** : qui garantissent le respect de la réglementation (RGPD, DGA, DA, etc.) ainsi que des attentes des utilisateurs traduites dans des labels exprimant des niveaux de confiance (de l'auto-description à des audits réalisés par des tiers)

“Les services de fédération facilitent l'émergence et le fonctionnement des espaces de données. Les espaces de données sont un concept d'intégration de données qui ne nécessite pas de consolidation des données en un seul endroit central, ni l'utilisation d'un seul schéma de base de données. En revanche, l'intégration des données est réalisée au niveau sémantique, et des redondances de données sont possibles.”⁵⁸

L'association Gaia-X AISBL met à disposition de ses membres et utilisateurs un document d'architecture générale⁵⁹ pour les *data spaces*, permettant de coordonner les développements de l'ensemble des acteurs.

3.3. BÉNÉFICES ET CAS D'USAGE

Les *data spaces* représentent une opportunité majeure de développement des écosystèmes numériques pour l'Europe et pour la France. **Les *data spaces* vont fluidifier les usages et favoriser l'innovation pour l'ensemble des acteurs de l'économie :**

- **Les acteurs privés** pourront (re)prendre le contrôle des canaux de distribution numériques face aux grandes plateformes au-delà de l'Open Data, augmenter leur visibilité et améliorer leur référencement et celui de leurs offres, permettre une expérience utilisateur fluide de bout en bout, favoriser une plus grande interopérabilité et des cas d'usage croisés innovants entre de multiples services numériques, certifier des données, bénéficier d'une redistribution équitable de la valeur (incluant la longue traîne des petits acteurs) et d'analyses de données.
- **Les acteurs publics** pourront offrir une expérience utilisateur fluide et augmentée pour les citoyens, associer des services privés à des services aux citoyens, réduire la charge administrative (le "Dites-le nous une fois"⁶⁰ évitant la ressaisie de données), certifier des données, s'appuyer sur les services d'identité numérique d'État, bénéficier de statistiques B2G (Business to Government) complètes, participer et contribuer au "Data Altruisme" (notion introduite dans le Data Governance Act et offrant un cadre pour l'exploitation et le partage des données dans un cadre altruiste, et non plus commercial ou contraint).
- **Les particuliers (utilisateurs de services numériques, citoyens)** bénéficieront d'une expérience utilisateur fluide indépendante des grandes plateformes, d'un haut degré de personnalisation (grâce à une IA digne de confiance), de moyens de contrôle de la circulation de leurs données (sur la base du RGPD) et de services alliant le public et le privé de façon fluide en fonction de leurs besoins, de moyens de s'impliquer dans

⁵⁸ La Revue européenne du droit, Hubert Tardieu & Boris Otto, décembre 2021, n°3

⁵⁹ https://www.gaiax.es/sites/default/files/2022-01/Gaia-X_Architecture_Document_2112.pdf

⁶⁰ <https://www.numerique.gouv.fr/services/guichet-dites-le-nous-une-fois/>

la gouvernance des applications utilisant leurs données.

A l'origine les projets de data spaces, envisagés dans le cadre des travaux du Fraunhofer ayant mené à la création de Gaia-X, étaient plutôt de nature industrielle. Ils s'inscrivaient dans le cadre de la stratégie industrielle allemande 4.0, cette dimension est toujours présente dans certains projets de data spaces en cours de construction.

L'une des autres promesse des data spaces est l'amélioration des services au citoyen, dans la droite lignée du projet précurseur X-Road⁶², une infrastructure d'échange de données (Data Exchange Layer) lancée en Estonie en 2001, qui avait permis à l'Estonie de digitaliser 99% de ses services publics.

Enfin les data spaces en construction investissent tous les secteurs de l'économie, y compris les secteurs plus orientés vers les consommateurs finaux, incluant la question du partage des données personnelles, et donc de la prise en compte du RGPD. Ces sujets sont portés par exemple par l'ONG MyData Global⁶³, lancée en Finlande en 2018 pour promouvoir une vision des données centrée sur l'humain.

La France pourra s'associer à des acteurs plus matures sur certains sujets au niveau européen, et pourra également montrer la voie, sur d'autres, en s'appuyant sur ses leaders dans marchés clés (ex : tourisme, éducation, agriculture et agro-alimentaire, grande distribution, santé, beauté, luxe). Voici quelques exemples de cas d'usage par marché :

GREEN DEAL	Mesure de l'impact carbone pour tous les secteurs dans le cadre de l'objectif Net Zero
EDUCATION ET EMPLOI	Réunir toutes les données de compétence, de personnalité et de préférences d'une personne pour l'aider dans sa formation tout au long de la vie avec des recommandations de formateur et d'emploi contextualisées et pertinentes en fonction du territoire
TOURISME ET MOBILITÉ	Parcours personnalisés, tourisme augmenté, handicap & accessibilité, alternatives vertes et subventions de mobilités douces, véhicule connecté
ADMINISTRATION	'Dites-le-nous une fois', utilisation des données administratives dans les services de la sphère privée (finance, assurance, santé, etc.), service au citoyen augmenté
SANTÉ	Carnet médical numérique, simplification administrative, données pour la recherche, médecine 5P (personnalisée, préventive, prédictive, participative, par preuves), nouveaux dispositifs médicaux (IoT), digital twinhealth
ÉNERGIE	Suivi des consommations, recharge véhicules électriques, flexibilité des réseaux, mix énergétique, renouvelables, recherche
AGRICULTURE ET AGRO-ALIMENTAIRE	Traçabilité de la ferme à l'assiette, bien manger, alimentation durable, gestion de l'eau, gestion des sols, données génétiques cheptel, recherche
SMART MANUFACTURING	Réutilisation de pièces, circulation de la donnée entre constructeurs et fournisseurs, décarbonation de l'industrie
SMARTBUILDING	Captation et analyse de données pour une meilleure gestion environnementale

⁶² <https://x-road.global/>

⁶³ <https://www.mydata.org/>

FINANCE	KYC (Know You Customer), détection des fraudes, gestion de budget, démarches administratives facilitées, aide à la réalisation de projets (immo, études, voiture,...)
CULTURE ET MÉDIAS	Détection de <i>fake news</i> , traçabilité des contenus, offres mutualisées presse, Métavers, profil décentralisé pour les réseaux sociaux
TERRITOIRES DIGITAUX	Smart City, <i>data spaces</i> régionaux, digital twins pour smart cities
SPACE	Exploitation des données d'observation de la terre pour l'environnement, l'agriculture et la sécurité, recherche, digital twin earth
ACTIVITÉS DE LA MER	Permaculture, données marines pour l'environnement, navigation, recherche, digital twin ocean, digital twin earth

CAS D'USAGE DES DATA SPACES PAR MARCHÉ

De nombreux cas d'usage, envisagés ou en cours de développement, ont été décrits de manière détaillée, et sont publiés dans les positions papers de Gaia-X⁶⁴. Les groupes de travail (GT) du Hub France de Gaia-X couvrent déjà la quasi-totalité de ces sujets.

3.4. LES DIFFÉRENTS TYPES D'ACTEURS DES DATA SPACES

Les *data spaces* sont composés d'acteurs de natures très variées, que l'on peut regrouper en 4 catégories :

- **Les institutionnels** (publics et parapublics, associations et fédérations/syndicats, régulateur, etc.). Ils jouent, ou peuvent jouer, un rôle essentiel dans de nombreux *data spaces* à différents titres. Par exemple :
 - L'État ou les collectivités sont producteurs ou consommateurs de données pour de nombreux cas d'usage de partage de données, notamment via la politique du 'Dites-le-nous-une-fois' et les feuilles de route ministérielles des données, des algorithmes et des codes sources.
 - Les enjeux de régulation sectoriels (parfois nationaux) ont un impact direct sur le développement et la gouvernance des *data spaces*, par exemple la loi mobilités⁶⁵ et son impact sur le data space mobilité.
 - Sur la question du partage des données personnelles, la CNIL (Commission nationale de l'informatique et des libertés) peut mettre en place des "bacs à sable réglementaires"⁶⁶ dédiés au partage de données.
 - Pour certains secteurs, les autorités régulatrices, par exemple la CRE (Commission de Régulation de l'Énergie) pour le data space Énergie, devraient jouer un rôle majeur dans la gouvernance.
 - La commande publique peut permettre le développement des *data spaces* et fixe un cadre pour leur développement pouvant aller jusqu'à imposer ou influencer certains choix techniques.
 - Certains cas d'usages vont s'appuyer sur MonFranceConnect⁶⁷ en cours d'expérimentation et l'évolution de l'identité numérique avec la mise en place du e-wallet européen et la norme eIDAS V2⁶⁸ dans le cadre du programme France Identité⁶⁹,

⁶⁴ <https://www.data.gouv.fr/fr/datasets/feuilles-de-route-ministerielles-sur-la-politique-de-la-donnee-des-algorithmes-et-des-codes-sources/>

⁶⁵ <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000039666574/>

⁶⁶ 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.

⁶⁷ <https://www.numerique.gouv.fr/espace-presse/lancement-de-lexperimentation-de-mon-franceconnect-letat-accelere-la-simplification-des-demarches-administratives/>

renforçant la confiance dans les échanges de données personnelles.

- **Les grands groupes** dans tous les secteurs sont de grands producteurs de données, qui pourraient devenir les moteurs de leurs écosystèmes respectifs :
 - Les grands groupes sont les leaders de verticales et doivent être des locomotives pour favoriser la coopération. Ils sont déjà regroupés dans des logiques de filière avec l'écosystème des PME (GIFEN⁷⁰, GiFAS⁷¹,...).
 - Souvent internationaux, les grands groupes auraient tout intérêt à s'investir plus franchement sur les *data spaces*, s'ouvrant ainsi à minima le marché européen.
 - Les *hyperscalers* n'apportant pas toujours de garanties substantielles de protection, beaucoup de données des grands groupes ne peuvent pas être hébergées sur des offres de cloud public. Nous estimons également qu'entre 20 % et 25 %⁷² des utilisateurs potentiels d'une suite collaborative SaaS, telle que Microsoft 365 ou G Suite de Google, ont besoin d'une solution de confiance et pourraient migrer vers un service hébergé sur une infrastructure de cloud et data (voire IA) de confiance. Si par ailleurs une telle suite collaborative s'avérait à l'état de l'art dans la durée, et si son coût d'usage n'était pas supérieur de 10 % à 15 % au coût de l'offre de référence du marché, alors plusieurs entreprises pourraient la choisir pour un nombre plus important d'utilisateurs, voire pour la totalité d'entre eux.
- **Les PME et startups innovantes :**
 - Les startups sont potentiellement de grandes réutilisatrices de données. Celles de la FrenchTech en particulier pourraient, si elles avaient plus accès aux données, proposer de nombreux cas d'usages innovants dans tous les secteurs (Fintech, Edtech, Healthtech, etc.).
 - Ce sont aussi des fournisseurs potentiels pour de multiples composantes essentielles de l'infrastructure des *data spaces*. En accompagnant le développement de startups de l'infrastructure des *data spaces* (ce que nous appelons "InfraTech"), la France pourrait faire naître les champions de demain qui auraient alors directement accès à un marché unique européen grâce aux standards d'interopérabilité, et ne se cantonneraient plus à leur marché national.
 - Les PME de manière générale doivent également pouvoir exploiter leurs données ou les valoriser, si elles veulent rester compétitives. Les *data spaces* contribueront directement à la numérisation du tissu des PME en offrant des services numériques partagés et inclusifs.
 - Pour les TPE la question des ressources est centrale, les services numériques partagés et inclusifs permet de mutualiser ces ressources qu'elles soient techniques (ex : data scientist en temps partagé) ou organisationnelles. Impliquer les TPE permet également d'apporter des données très locales, auxquelles les grands groupes n'ont pas accès, offrant ainsi des services beaucoup plus pertinents, et à plus forte valeur ajoutée. L'un des enjeux de Gaia-X est de développer des initiatives locales, et de ne pas tout concentrer "en haut", y compris via des modèles économiques soutenables pour ceux qui n'ont pas les moyens financiers des grands groupes.
- **Les acteurs de la recherche :**
 - Les chercheurs développant des services et modèles innovants pourront accéder à davantage de données pour leurs recherches via les *data spaces*.
 - Ils seront également des acteurs majeurs de la construction de l'infrastructure des

⁶⁸ <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eIDAS-Node+version+2.0>

⁶⁹ <https://france-identite.gouv.fr/en-savoir-plus/de-quoi-s-agit-il/>

⁷⁰ Groupement des Industriels Français de l'Énergie Nucléaire

⁷¹ Groupement des industries Françaises Aéronautiques et Spatiales

⁷² chiffres issus d'un groupe de travail restreint *cloud de confiance*, piloté par Vincent Niebel, DSI du groupe EDF.

data spaces et assureront un développement interdisciplinaire respectant l'état de l'art (technique, éthique, légal, business, sciences sociales, enjeu sociétal, environnemental,...).

Chacun de ces acteurs peut exercer l'une ou plusieurs de ces différentes fonctions/ rôles pour les *data spaces* :

- **Les utilisateurs de *data spaces* sont les acteurs métier** (ex : EDF⁷³ pour le *data space* Énergie), parties prenantes des *data spaces*, qui souhaitent mettre à disposition et/ou consommer des données, de manière contrôlée, et au travers d'un modèle économique soutenable et prenant en compte toutes les parties prenantes.
- **Les fournisseurs d'infrastructure (InfraTech)** sont les startups/scale-up (ex : Aleia, Archipels, Cleyrop, Dawex, OKP4, Onecub, OpenDataSoft, OVH, Inokufu, Murmuration, Seraphin.legal, Smart B, Startin'Blox, Visions, etc.) et grands groupes (ex : 3DS Outscale, Atos, Cap Gemini, Digiposte/Docaposte, Orange Business Services, Sopra Steria, etc.), capables de fournir les composants logiciels, open source ou non, de l'infrastructure des *data spaces* (gestion des identités, consentements, API, gestion légale, business, blockchain/Web3, interopérabilité, etc.).
- **Les exploitants de *data space*** sont les entités chargées de la gestion des *data spaces* (ex : Agdatahub⁷⁴ pour le *data space* agriculture). Certains *data spaces* peuvent fonctionner sans incarnation formelle, en s'appuyant uniquement sur des standards d'interopérabilité au sein d'une filière par exemple, d'autres *data spaces* seront exploités par des organisations (associations, sociétés privées, GIE, GIP, coopératives, etc.). Les exploitants peuvent être de nature publique ou opérer via des délégations de services publics (ex : sur les sujets régaliens), publique/privée, ou purement privée (SAS, SA,...). Ils s'assurent que l'ensemble des acteurs du *data space* puissent participer à la gouvernance commune. Le modèle économique qui sera mis en place par ces exploitants sera un facteur clé de durabilité. Ils devront générer leurs propres ressources, et ne pas être uniquement dépendants de subventions. Les modèles peuvent se baser sur la vente de services (accompagnement, développements spécifiques), ou sur une rémunération sur les transactions pour les données (rémunération de toutes les parties prenantes).
- **Les administrateurs de la gouvernance des Communs Numériques**, sur lesquels reposent l'infrastructure des *data spaces*, sont la plupart du temps des associations telles que Gaia-X sur les sujets généraux, ou MyData⁷⁵ sur la question particulière du partage des données personnelles ou encore certains *data spaces*, ou groupements de *data spaces*, eux-mêmes sur les questions de communs numériques sectoriels (ex: Prometheus-X pour les communs des données d'éducation). Elles garantissent que les *data space* reposent sur des spécifications, et du code open source, afin qu'aucun acteur ne verrouille la capacité à se connecter à un *data space*. Ces organisations peuvent également proposer des labels de confiance (ex : Label Gaia-X⁷⁶, certification MyData Operator⁷⁷, etc.).

⁷³ <https://www.gaia-x.eu/news/member-story-electricite-de-france>

⁷⁴ <https://agdatahub.eu/>

⁷⁵ <https://www.mydata.org/>

⁷⁶ https://gaia-x.eu/wp-content/uploads/files/2022-02/Labeling_Criteria_Whitepaper_v07.pdf

⁷⁷ <https://oldwww.mydata.org/mydata-operators/>



RÉGLER LES PROBLÈMES
DE GOUVERNANCE DU
PARTAGE DES DONNÉES,
EST PLUS SIMPLE QUE
COMBLER NOTRE RETARD
TECHNOLOGIQUE.

3.5. UN BESOIN FORT DE GOUVERNANCE

La gouvernance au cœur des *data spaces*

La variété des acteurs, leur nombre, ainsi que leurs différents enjeux et niveaux de maturité, impliquent d'abord un effort de coordination conséquent pour homogénéiser les pratiques. Au-delà de la coordination, les acteurs d'un *data space* doivent ensuite être en mesure de s'accorder sur des règles communes, et de prendre des décisions claires, si aucun consensus n'est trouvé.

Les acteurs du marché, autant les startups que les TPE/PME ou les grands groupes, seront la clé du succès des *data spaces*, car ce sont eux qui construiront l'infrastructure et produiront la plupart des services de demain. Ils doivent donc s'approprier ces nouvelles logiques de "coopétition". Cependant **nous pensons que seul un État stratège, moteur, conscient des enjeux et jouant pleinement son rôle aux côtés de tous ses partenaires européens (à commencer par l'Allemagne), permettra aux projets de data spaces d'émerger et de se développer.** En effet, par la commande publique et l'investissement, l'État a la capacité d'assurer le lancement des *data spaces* sur la quasi-totalité des secteurs, et de structurer le marché là où **les acteurs privés ont souvent du mal avec les approches mutualisées...** Par ces mêmes moyens l'État est aussi en mesure d'aligner les développements sur une doctrine commune. Enfin, un grand nombre de données, ainsi que certaines décisions de gouvernance, relèvent de la prérogative des États, qui sont par ailleurs seuls légitimes à garantir la protection des citoyens, la confiance ne pouvant être l'exclusive des intérêts privés.

Les enjeux par type d'acteur

Chaque type d'acteur présente des problématiques qui lui sont propres concernant le partage de données :

- **Les institutionnels** développent rarement une vision globale des enjeux (ex : vision de numérisation centrée sur l'administration plutôt que de la filière elle-même, difficulté à coordonner l'action entre les ministères) et peuvent éprouver certaines difficultés à s'adapter aux problématiques des autres acteurs. Ils ont également une adoption de l'innovation plus lente et sont souvent réticents au partage de données, tout comme à la collaboration avec les acteurs privés.
- **Les grands groupes** expriment des réserves stratégiques importantes vis à vis du partage des données : standards peu clairs, craintes réglementaires, peur d'ouvrir ses données à la concurrence, comportement de "gatekeeper", difficulté à se projeter sur le long terme (retour sur investissement incertain), difficulté à visualiser les usages.
- **Les PME et startups innovantes** manquent de moyens (développements techniques coûteux, 'temps' des startups versus 'temps' des grands groupes et institutionnels) et éprouvent souvent des difficultés à s'impliquer dans l'effort de standardisation, ou à s'insérer dans les appels d'offres de consortiums.

Les besoins de coordination intra et inter *data spaces*

Un effort de coordination doit avoir lieu au sein de chaque data space, où il est nécessaire de faciliter le dialogue entre les différents types d'acteurs qu'il est parfois complexe de faire cohabiter, d'arbitrer les conflits potentiels, d'animer les écosystèmes, de coordonner les

développements, d'identifier et de lever les barrières à l'entrée, et de construire des modèles économiques communs et pérennes.

L'effort doit également avoir lieu entre les *data spaces* afin de permettre la mutualisation des développements (standards, Communs Numériques, etc.) et l'interopérabilité multi-sectorielle et trans-nationale.

C'est bien l'effort combiné de l'ensemble des acteurs du, ou des, *data space* qui permettra d'offrir une véritable valeur ajoutée aux utilisateurs finaux, ainsi qu'aux fournisseurs de données eux-mêmes.

L'État français doit selon nous aujourd'hui favoriser le développement d'une vision d'ensemble (public/privé) sur la question des *data spaces* et du partage de données afin d'aider tous les acteurs à se coordonner.

L'EXEMPLE D'ALBERTO PALOMO-LOZANO CHIEF DATA OFFICER DE L'ESPAGNE

Le Gouvernement espagnol a créé il y a un peu plus de 2 ans cette fonction clé, qui dépend du Vice-Premier Ministre et du secrétaire d'État à la numérisation et à l'intelligence artificielle, et dont l'objectif est d'optimiser l'utilisation des données.

L'une de ses missions est de veiller à ce que le développement de l'IA se fasse de manière "alignée et coordonnée" entre les différentes branches du gouvernement et les entreprises privées, notamment celles qui finissent par participer à l'un des projets clés réalisés avec des fonds européens.

Une autre de ses tâches consiste à **promouvoir la mise en œuvre de l'architecture de partage de données développée au sein de Gaia-X et les relations avec l'industrie en termes de partage de données et de services cloud**. En collaboration avec le centre national Gaia-X en Espagne, le bureau des données soutient le déploiement d'espaces de données sectoriels, en mettant l'accent sur les secteurs du tourisme, de la santé, de l'agroalimentaire et de la mobilité durable.

M. Palomo n'est pas un haut fonctionnaire. Il est titulaire d'un doctorat en physique théorique, il termine actuellement un certificat en gestion, technologie et innovation au MIT. Il est expert en Data Analytics et en définition de stratégies de données, possède une grande expérience de la gestion d'équipes de data scientists au niveau international pour générer des gains d'efficacité dans les processus métier, et a travaillé à la conception de matériel pour optimiser la phase d'entraînement des réseaux neuronaux basés sur l'intelligence artificielle.



Exemples de défis de coordination

Plusieurs *data spaces* présentent des défis intra ou inter-sectoriels pour lesquels l'implication de l'État est un prérequis indispensable :

- **L'exemple du *data space Green Deal*⁷⁸** : le *Green Deal* européen a l'ambition de relever les défis liés au climat et à l'environnement. Il intègre une stratégie de croissance tout en visant à protéger, conserver et améliorer le capital naturel de l'UE, et à protéger la santé et le bien-être des citoyens contre les risques et les impacts liés à l'environnement. En permettant la circulation des données liées à l'impact environnemental, le *data space Green Deal* se positionne comme un outil permettant de repenser les politiques d'approvisionnement en énergie verte dans toute l'économie (industrie, production & consommation, grandes infrastructures, transport, alimentation, agriculture, construction,

⁷⁸ <https://www.cigref.fr/wp/wp-content/uploads/2021/10/French-Gaia-X-Hub-Position-Paper-GT-Green-deal-long-version-Vf.pdf>

fiscalité, prestations sociales,...). Il permettra également d'augmenter la protection et la restauration des écosystèmes naturels, de favoriser l'utilisation durable des ressources, et d'améliorer la santé des citoyens. Par nature trans-sectoriel et au croisement d'enjeux de natures privées (acteurs de l'énergie, commerçants, industriels, promoteurs immobiliers,...) et publiques (villes, régions, administration fiscale,...), le data space Green Deal est l'affaire de tous les acteurs de l'économie, sans pour autant être le cœur d'activité de la plupart d'entre eux. Concernant le secteur public, on remarque que de nombreux ministères devraient être partie prenante d'une démarche commune. Seul l'État est actuellement en mesure d'aligner les multiples intérêts publics et privés de l'ensemble de ces acteurs, par la vision, la planification, l'investissement, ainsi que par la commande publique.

- **L'exemple du data space Éducation & Compétences**⁷⁹ : Le data space éducation doit permettre au citoyen de bénéficier d'un CV dynamique, qui le suivra tout au long de son parcours professionnel. Ce CV est construit à partir de données provenant du ministère de l'éducation nationale, de ses multiples établissements scolaires, de ses organismes de formation, de Pôle Emploi, de Mon Compte Formation⁸⁰, des grandes entreprises dans lesquelles il a évolué, de startups de la Edtech qui l'auront aidé à s'orienter, etc. Des initiatives telles que le programme Erasmus⁸¹, facilitant la mobilité des étudiants en Europe, appellent également naturellement à joindre les initiatives à l'échelle européenne. Sans la participation active et centrale de l'État, tant au niveau de la gouvernance de l'accès aux données que du financement de l'infrastructure régaliennne, le data space éducation ne peut pas se développer.

Les premiers retours d'expérience

Nous devons nous appuyer sur les premiers retours d'expérience (éducation, mobilité, énergie, agriculture, etc.) pour être en mesure de généraliser la création des data spaces à tous les secteurs et territoires.

A de rares exceptions près, les premiers projets de data spaces font état d'une vraie difficulté à financer des initiatives de place publiques-privées de consortium, par nature complexes. L'inadéquation des modalités de certains appels à projet publics existants, par rapport aux enjeux inhérents aux projets de data spaces, rend la tâche particulièrement ardue pour les porteurs de projets, qui par ailleurs sont souvent des entrepreneurs rassemblant des écosystèmes de dimension régionale ou nationale, avec des moyens très limités.

Les difficultés les plus récurrentes sont les suivantes :

- temps longs d'instruction des appels à projet dépassant parfois l'année, ce qui démotive les partenaires,
- part importante d'avances remboursables, y compris pour financer des Communs Numériques ou des tâches qui peuvent être considérées comme régaliennes,
- cadre trop rigide ne prenant pas en compte l'état de l'art du marché naissant des data spaces,
- difficultés à valoriser les apports des parties prenantes, en particulier pour les startups,
- difficultés à identifier les appels à projet au niveau européen (au regard des montants concernés il faudrait avoir un support pour détecter ces financements)
- appels à projet sans lien direct avec l'échelon de marché européen ou le programme de financement Digital Europe (pourtant cadre naturel des projets de data spaces),

⁷⁹ <https://prometheus-x.org/pages/gaia-x?locale=fr>

⁸⁰ <https://www.mon-compte-formation.fr/>

⁸¹ <https://www.education.gouv.fr/le-programme-erasmus-9890>

- appels à projet ou à manifestation d'intérêt poussant les acteurs à recréer chacun de leur côté une infrastructure de partage de données en propre. Ce qui d'une part crée une concurrence injustifiée entre les uns et les autres, et surtout leste lourdement le coût de leur projet (d'où l'ambition de cette note de mutualiser les investissements pour une infrastructure commune qui permettra d'alléger l'assiette des *data space* et évitera aussi à l'Etat d'avoir à financer plusieurs fois les mêmes éléments d'infrastructure)

Certains appels à projets, liés aux infrastructures logicielles de partage de données, s'inscrivent en revanche dès le départ dans une démarche européenne, affichent une volonté forte de réutilisabilité des développements entre les projets, et incluent de multiples Ministères. C'est le cas par exemple de l'appel à projet « territoires intelligents et durables » de France 2030⁸², associé à la démarche européenne Living-in.eu⁸³ et à Eurocities⁸⁴.

On constate que la totalité des projets de *data spaces* en cours de construction en France, et en Europe d'ailleurs, ont des besoins d'infrastructure logicielle de partage de données communs (gestion des consentements, de l'identité, des modèles légaux et business, des API,...). **Pourtant les porteurs de projets éprouvent de grandes difficultés à mettre en commun leurs développements**, y compris sur des couches basses d'infrastructure, car ils sont mis en concurrence entre eux sur les appels à projet publics. **Les acteurs métier de leur côté** (éducation, agriculture, tourisme, mobilité,...) **n'affichent pas naturellement une volonté de financer les couches communes** (ni de manière indépendante, ni en mutualisant leurs efforts) et préfèrent orienter leurs financements vers les cas d'usage et les applications métier. L'incapacité de nos acteurs à se rassembler, et à mettre en commun leurs efforts, favorise la conquête de marché des *hyperscalers* américains, tant sur le plan des solutions packagées Cloud-Data-IA "bout en bout" que sur les questions de communs numériques. Pourtant nous disposons déjà en Europe de toutes les compétences nécessaires, et des moyens, pour couvrir l'ensemble de la chaîne de valeur.

⁸² Cet AAP a vocation à soutenir les collectivités territoriales, syndicats mixtes et intercommunaux, ayant pour ambition d'apporter grâce à la donnée une solution à un problème de pilotage de politiques publiques ou d'améliorer la gestion d'un service aux usagers.

⁸³ <https://living-in.eu/>

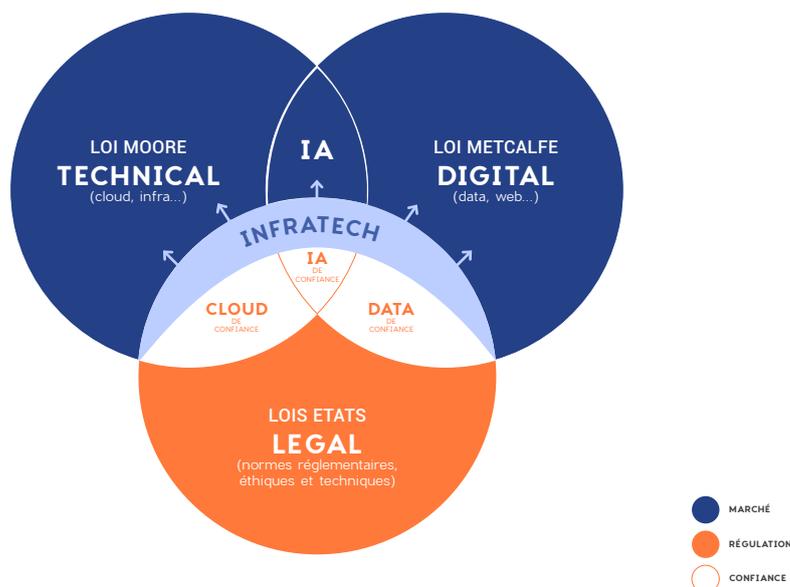
⁸⁴ <https://eurocities.eu/>



POUR QUE LA STARTUP
NATION NE DEVIENNE
UN MIRAGE, LA FRANCE
DOIT AUSSI DEVENIR
UNE INFRASTRUCTURE
NATION.

IV. DÉVELOPPER UNE INFRATECH DU PARTAGE DE DONNÉES

ÉCOSYSTÈME DE CONFIANCE (CLOUD + DATA + IA) = AUTONOMIE STRATÉGIQUE



Notre conviction est que la France ne doit pas se contenter de devenir une startup nation, car plus elle développe des startups métier, plus paradoxalement elle amplifie ses dépendances (ces startups étant de grandes consommatrices des services techniques et marketing offerts par les *hyperscalers* extra-européens). Pour que ce magnifique tissu économique de la tech soit protégé, pour que les écosystèmes deviennent véritablement systémiques, et pour que chaque euro investi dans ces startups ne bénéficie pas uniquement aux *hyperscalers*, la France **doit devenir également une *infrastructure nation***⁸⁵.

C'est la condition de notre autonomie stratégique, et ce serait surtout un legs majeur aux générations actuelles et futures de citoyens. Comme le nucléaire avant lui avec le plan Messmer, le numérique mérite également une stratégie étatique. D'autant plus qu'il s'agit ici d'une "soft infrastructure immatérielle" qui ne coûterait que quelques "centaines" de millions d'euros, presque un "détail" si nous les comparons au 100 milliards investis dans les infrastructures nucléaires dans les années 70. **La France doit selon nous contribuer à la création de ce que nous appelons "l'écosystème InfraTech européen" : un écosystème d'acteurs européens spécialisés dans le logiciel, souvent start-ups et PME, qui se fédèrent en vue de proposer une offre de bout en bout qualifiée d'infrastructure logicielle compétitive pour le cloud, la data ou l'IA.** Elles portent intrinsèquement les valeurs et la réglementation européenne, des normes harmonisées ou des standards partagés. Elles s'appuient sur des communs numériques qu'il s'agira de valoriser collectivement sur le marché dans une stratégie horizontale irriguant ainsi l'ensemble des secteurs stratégiques pour l'Europe (emploi, éducation, santé, finance, tourisme, mobilité, administration, etc.). Elles requièrent le développement d'environnements logiciels fédérateurs (ou infrastructure logicielle) pour garantir la couverture de l'ensemble de la chaîne de valeur dans le domaine concerné,

⁸⁵ Tariq Krim

assurer l'interopérabilité et permettre le partage de valeurs au sein de l'écosystème. Enfin, elle nécessite une déclinaison « métiers » dans le cadre de « data spaces » afin de garantir son déploiement et son adoption par les secteurs applicatifs.

4.1. DÉCENTRALISER L'INFRASTRUCTURE PAR UNE SURCOUCHE D'INTEROPÉRABILITÉ

L'Europe doit être en mesure de définir et de proposer une nouvelle forme d'architecture pour le partage de données, qui corresponde à ses propres objectifs de souveraineté et de préservation des droits fondamentaux. La définition du Cloud proposée par les grands *hyperscaler*, soit une offre packagée et individuelle qui couvre toute la chaîne de valeur Cloud-Data-IA, n'est pas forcément celle que nous retiendrons. "Un juste équilibre entre l'intérêt du détenteur des données individuelles et l'intérêt légitime de la communauté à utiliser les données existantes (par exemple, à des fins de soins de santé, pas seulement en temps de pandémie actuelle) est au centre du débat sur la manière de concevoir l'économie européenne des données. L'équilibre entre les intérêts des individus et les intérêts communs conduit à une conception d'infrastructure différente de ce qui est disponible dans les offres actuelles des fournisseurs de plateformes privées américains ou chinois."⁸⁶

L'approche européenne est plus décentralisée et "démocratique" que les approches des hyperscalers, même si les États-Unis commencent déjà à mettre en place des infrastructures partagées. "Toutes les parties prenantes s'accordent à dire que l'infrastructure de données et de logiciels de l'Europe doit empêcher l'émergence d'une centralisation du pouvoir en matière de données. C'est essentiel dans la mesure où cela nécessite d'empêcher que les effets de réseau conduisent à une situation où « le gagnant prend tout »"⁸⁷.

"Les infrastructures européennes sont conçues comme une fédération. Elles ne nécessitent pas de fonctionnalités centrales de stockage, de traitement ou de distribution des données. Au lieu de cela, l'infrastructure est formée de nœuds distribués dans un réseau de données et de services. Les nœuds peuvent fournir ou utiliser des ressources telles que des données. Les fournisseurs restent indépendants et libres en ce qui concerne leurs données et autres ressources ; le réseau est ouvert et non discriminatoire et enfin les utilisateurs et les fournisseurs se font mutuellement confiance."⁸⁸

⁸⁶ La Revue européenne du droit, Boris Otto & Hubert Tardieu, décembre 2021, n°3

^{87/88} La Revue européenne du droit, Boris Otto & Hubert Tardieu, décembre 2021, n°3

Développer l'infrastructure (ou InfraTech) du partage de données, et des *data spaces* de manière plus générale, consiste en la création d'une surcouche d'interopérabilité Cloud-Data-IA qui pourra s'intégrer ou s'ajouter aux offres cloud existantes (celles des *hyperscalers*, comme celles des autres acteurs), afin de créer un lien virtuel et décentralisé entre les services métier d'une multitude d'acteurs.

CLOUD INFRASTRUCTURE



Des travaux d'interopérabilité et d'horizontalisation sont en cours sur l'ensemble des 3 composantes de l'écosystème de confiance Cloud-Data-IA au travers d'initiatives telles que Gaia-X, ou le Grand Défi IA français.

	CLOUD	DATA	IA
ÉCOSYSTÈMES MÉTIER	Utilisateurs de services Cloud	Utilisateurs des <i>data spaces</i> (ex : EDF, Accor Hôtel, APHP, startups sectorielles, Ministères, etc.)	Utilisateurs d'IA de confiance
ÉCOSYSTÈMES D'INFRASTRUCTURE	Cloud providers (ex : OVH, Microsoft Azur, etc.)	Acteurs infrastructurants du data-sharing (ex : e-Wallet, data sharing intermediaries, fournisseur d'infrastructure de data spaces, etc.) (ex : OKP4, Onecub, Visions)	Acteurs infrastructurants de l'IA de confiance (ex : outils et méthodes développés dans le cadre du programme Confiance.ai)
COMMUNS NUMÉRIQUES	GXFS (Gaia-X)	GXFS (Gaia-X) (ex : Agdatahub, Prometheus-X, Themis-X)	Communs Confiance.ai
STANDARDS	Gaia-X (Label), SecNumCloud (Fr), C5 (All), ENISA à venir	Gaia-X, IDSA, BDVA, MyData, etc.	Pilier #3 Normes du Grand Défi IA (SGPI)

COMPOSANTES DE L'ÉCOSYSTÈME DE CONFIANCE

Les **data spaces** sont des **meta-plateformes de services régies par des règles d'interopérabilité**, qui se traduisent par des API socles (conformes par exemple aux standards Gaia-X). Ces meta-plateformes sont un assemblage intelligent et complet de briques créées par un ensemble d'acteurs qui s'entendent entre eux via :

- l'interopérabilité des briques au sein d'un data space, ce qui permet de créer un effet de réseau puissant, et de mettre en œuvre une offre de service large et complète ;
- l'alliance entre les data spaces (via Gaia-X), qui permet de générer une masse critique à l'échelle de l'UE, et ainsi de fédérer tous les acteurs européens autour d'une même démarche.

Les **data spaces fédèrent un ensemble de services proposés par des consortiums, fournis sous forme de plateforme "packagée" et hébergée. On peut parler de Platform as a service (PaaS).**

Au sein du bloc Portail d'un data space, on retrouve par exemple :

- 1 brique Interface web qui est le "Front-End" du Portail
- 1 brique Onboarding qui permet de rejoindre l'écosystème
- 1 brique Gouvernance (acceptation des demandes d'adhésion, accord sur les termes et conditions, l'organisation, l'opérationnel et les réglementations type RGPD, SLAs)
- 1 brique Catalogue pour ajouter, cataloguer et utiliser des ressources
- 1 brique « Services d'Infrastructure » (services d'interopérabilité, de traçabilité, ou tout service permettant la circulation et la valorisation des données et la gouvernance de l'écosystème).

L'approche environnementale des data spaces sera un autre élément clé de leurs succès. Cette approche s'exprime dans le principe même des data spaces qui ne stockent pas (ou peu) de données, mais les cataloguent et mettent en œuvre des moyens de stockage éphémères, uniquement si nécessaire. **En ce sens les data spaces contribuent à une vision responsable et frugale de la circulation des données.** Par ailleurs, les principes de l'identité décentralisée (basés sur la Blockchain) et les portefeuilles (eWallet) qui conservent les différentes informations nécessaires au suivi des traitements et des transactions effectuées, sans la stocker dans sa totalité, contribuent à la minimisation des données stockées, en rendant chaque acteur responsable de ses données.

L'infrastructure des data spaces est "portable" et susceptible de changer d'environnement cloud à la demande. Les principes de « Communs Numérique » open source, et la volonté de livrer des composants avec une couche d'orchestration et des règles communes de documentation; permettent de simplifier leur déploiement, et donnent des garanties quant à la durabilité et la résilience de ces services.

De nombreux cas d'usage, envisagés ou en cours de développement, ont été décrits de manière détaillée, et sont publiés dans les positions papers de Gaia-X⁸⁹. Les groupes de travail (GT) du Hub France de Gaia-X couvrent déjà la quasi-totalité de ces sujets.

4.2. S'IMPLIQUER PLEINEMENT DANS L'EFFORT DE STANDARDISATION

Les principaux acteurs du cloud, les hyperscalers, n'ont pas forcément intérêt à voir émerger une fédération d'infrastructures interopérables qu'ils ne maîtriseraient pas, leurs business models étant basés sur une captation de la valeur verticalisée. Ils ont une très grande part de marché sur le Cloud et pourraient devenir les standards de facto des *data spaces* qui s'appuient sur ces clouds. Ils participent par ailleurs déjà à la création de composants open source pour les *data spaces*. Si nous prenons l'analogie du véhicule autonome, son système d'exploitation sera au final celui qui pourra fournir tous les services de manière "packagée" et deviendra *de facto* la plateforme de l'automobile. Si aucune règle n'existe autour de ce système d'exploitation, il peut fermer le marché à tout nouveau service. Un récent rapport de la Commission⁹⁰ a souligné que **la concurrence de géants des technologies, qui ne respectent pas toujours les règles et les valeurs fondamentales européennes et qui placent l'appropriation et la valorisation des données au cœur de leur stratégie, constitue un défi politique majeur pour l'Europe.**

Les hyperscalers s'organisent et défendent leurs intérêts business propres dans les initiatives de type Gaia-X, rendant parfois difficile le développement de règles claires et transparentes pour un marché plus juste. Par ailleurs l'Europe a cette complexité d'être composée d'États membres qui n'ont pas tous la même politique d'indépendance technologique. La France a toujours été pionnière sur ces sujets, **malheureusement nous sommes peu organisés et trop peu nombreux.** La mobilisation de certains acteurs de Gaia-X, notamment côté français (on peut mentionner ici le rôle moteur d'EDF) a malgré tout permis de poser les bases d'un système de labellisation Gaia-X, qui reprend les ambitions de souveraineté européenne. Le niveau le plus élevé de labellisation embarque par exemple l'immunité aux lois extra-territoriales, conformément aux critères établis par l'ANSSI, et ce malgré une opposition frontale des lobbyistes (pour certains européens malheureusement) représentant des intérêts non européens. Mais toutes ces actions restent très fragiles et reposent, comme nous l'avons dit, sur quelques bonnes volontés isolées, qui mériteraient un

⁸⁹ <https://projects.eclipse.org/projects/technology.dataspaceconnector>

⁹⁰ Digital sovereignty for Europe, EPRS Ideas Paper Towards a more resilient EU

appui, notamment juridique, à la hauteur des moyens de lobbying des hyperscalers américains.

Ce travail a un coût et personne à ce stade ne souhaite le financer. Quelques grands groupes ont pu dédier des ressources à la défense de leurs intérêts, mais cet effort est bien trop faible par rapport aux forces déployées par les *hyperscalers*. **Presque aucune PME ou startup innovante n'a les moyens de venir directement s'investir dans ces travaux.**

Dans un contexte où **Thierry Breton**, Commissaire européen au marché intérieur, **souhaite faire de l'Europe un chef de file de la norme et un "standard-setter" mondial, nous pensons qu'il est nécessaire de mutualiser l'action des acteurs français (institutionnels, grands groupes, startups, académiques) et de les aider à participer à l'effort européen de normalisation.**

4.3. CONSTRUIRE LES COMMUNS NUMÉRIQUES DU PARTAGE DE DONNÉES

LES COMMUNS NUMÉRIQUES

Les Communs Numériques s'inscrivent dans la réflexion plus générale sur les biens communs. Au croisement de la théorie économique et de la science politique, le thème a été popularisé par le travail d'Elinor Ostrom en 1990 puis son prix Nobel d'économie en 2009. Les biens communs désignent au départ des ressources physiques (rivière, forêt, pêche) pouvant être administrées collectivement.

Les Communs Numériques ont pour spécificité un coût de copie et de diffusion proche de zéro. Ce sont par exemple des codes sources, des données ou bases de données, des logiciels ou du contenu numérique (image/vidéo/son). À cela s'ajoutent les licences libres, qui organisent juridiquement l'accès, l'utilisation, la production, la modification, la diffusion et la gestion de ces ressources numériques.

Les Communs Numériques se distinguent des communs physiques par deux caractères principaux : ils sont non-exclusifs (pas de limite d'accès si quelqu'un utilise la ressource), et ils sont non-rivaux (l'usage de la ressource ne prive pas les autres usagers, elle reste disponible)



Les data spaces vont être construits sur des composants logiciels communs et gouvernés par une communauté d'acteurs, via les organisations impliquées dans la construction des *data spaces* : tous les aspects du partage de données liés à la confiance (consentement, contrats de partage de données, etc.), mais aussi la façon dont nous allons décrire les données ou ressources, la façon dont nous allons contrôler l'accès aux données, les identités des personnes et des organisations, les API qui vont permettre de faire transiter les données, etc.

Certains composants sont communs à tous les data spaces et ils sont stratégiques. L'Europe les appelle les « smart middleware ». Certains sont couverts par les services de fédération de Gaia-X (GXFS). Les architectures et spécifications sont définies dans Gaia-X (dans ses groupes de travail et ses comités qui nécessitent un investissement et un effort important).

D'autres composants peuvent être spécifiques à une filière (ex : les Communs des *data spaces* éducation européens comprenant par exemple des référentiels de compétences) et appellent à une coopération à l'échelle nationale ou continentale.

L'open source tient ici une part essentielle, il est une composante fondamentale de l'indépendance technologique et permet de garantir la multitude des acteurs. Ces phases d'émergence de solutions open source sont fragiles et ont besoin de soutiens financiers. Les *hyperscalers* l'ont parfaitement compris puisqu'ils sont capables de mettre à disposition de grands nombres de développeurs sur ces sujets. Par exemple, un acteur du cloud américain s'impliquant dans l'open source va mettre des moyens conséquents pour que ces composants ne se déploient que sur ses solutions, ou ne soient compatibles qu'avec elles.

Les premiers composants open source des *data spaces* vont arriver en 2022. **Il ne s'agit pas seulement de développer des composants, l'animation de la communauté open source va être déterminante** pour favoriser l'adoption, mais aussi le développement de tous les services qui s'appuieront sur ceux-ci, et la maintenance nécessaire pour les secteurs industriels les adoptant. L'initiative Simpl⁹¹ de la Commission européenne s'inscrit pleinement dans cette approche. Les acteurs français doivent impérativement s'impliquer fortement dans ces travaux européens, s'ils veulent exister plus tard dans les écosystèmes métier des *data spaces*.

Nous pensons que le niveau d'investissement dans les Communs Numériques du partage de données sera déterminant pour permettre aux acteurs français et européens de se développer sur le marché des *data spaces*. Nous devons continuer à identifier les besoins communs entre tous les *data spaces* en construction, et mutualiser les investissements. **Si nous n'agissons pas, chaque initiative métier de data space (santé hier, éducation demain,...) aura le choix entre redévelopper sa propre infrastructure, ce qui serait trop coûteux et freinerait l'interopérabilité, soit faire appel à des solutions non européennes.**

Ne pouvant se résigner à une telle impasse stratégique, nous appelons de nos vœux à une prise en main par la France de ces enjeux. C'est le rôle de l'État stratège de s'assurer que les infrastructures qui conditionnent et sous-tendent aux marchés d'avenir soient bien financées selon un plan national. A l'instar des coopératives agricoles en 1880 qui avaient été co-financées par l'État via le Crédit Agricole Mutuel et l'organisation des marchés⁹², ou bien le réseau ferré pour le transport ferroviaire avec le plan Freycinet, le marché de la donnée doit s'appuyer lui aussi sur une infrastructure dont le coût sera largement compensé par la création de valeur qu'elle permettra. Évidemment il n'est pas question d'aller jusqu'à un financement entièrement public pour une entreprise affectataire comme c'est le cas avec la SNCF, mais plutôt comme une organisation faîtière publique-privée du type GSMA dans les télécoms. Nous sommes aujourd'hui sur le numérique à un moment comparable à celui des télécoms au début des années 90, lorsque l'Europe avait fait le choix de la coopération en investissant sur des standards communs pour faire face au monopole de Motorola.

C'est pour cela que nous devons nous assurer que les marchés publics ayant pour vocation d'atteindre cet objectif soient bien en capacité d'offrir à nos écosystèmes les moyens de s'imposer. Car peser dans Gaia-X, ses Communs Numériques et standards aujourd'hui, c'est garantir demain notre autonomie stratégique. L'expérience récente ayant montré qu'il était impossible de faire porter de tels investissements par une seule filière, et ne pouvant nous résoudre à devoir choisir par dépit les solutions américaines ou chinoises, qui plus est sur des thématiques régaliennes d'intérêt général, nous proposons donc une autre voie basée sur la mutualisation et la coopération, tout en soutenant les projets déjà lancés et en attente de financements (GXFS, *data space* éducation, *data space* tourisme, etc).

Un tel effort serait écologiquement souhaitable et économiquement stratégique pour l'ensemble du territoire. Le tout pour l'équivalent du coût de quelques centaines de mètres de ligne à grande vitesse...

⁹¹ <https://digital-strategy.ec.europa.eu/en/news/simpl-cloud-edge-federations-and-data-spaces-made-simple>

⁹² *Emergence, développement et rôle des coopératives agricoles en France. Aperçus sur une histoire séculaire* Philippe Nicolas, 1988

4.4. DÉVELOPPER L'ÉCOSYSTÈME DES ACTEURS DE L'INFRA TECH

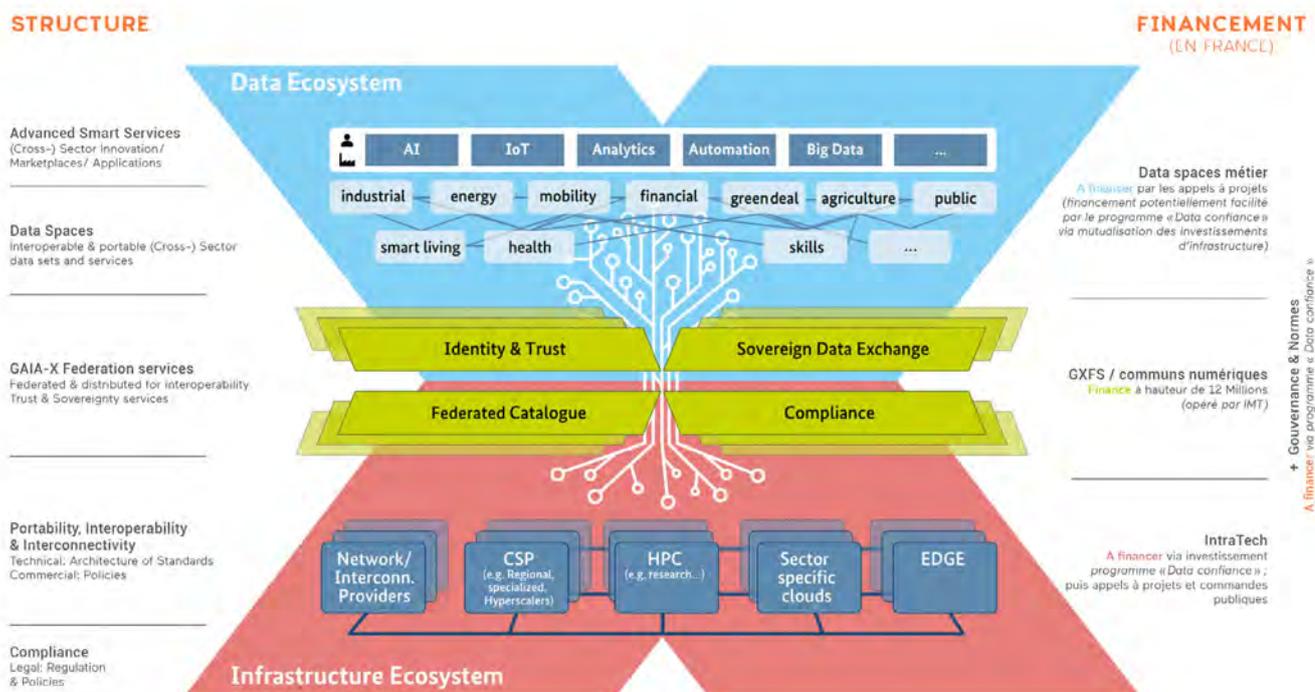
Face aux *hyperscalers* et dans le contexte de la stratégie des données européennes, les *data spaces* représentent une véritable opportunité pour la France et pour l'Europe de développer un écosystème d'acteurs de l'infrastructure du partage de données. Le RGPD par exemple n'a pas permis l'émergence d'un tel marché, les géants américains se sont emparés du sujet avec des moyens dont aucun acteur européen ne disposait, et de nouveaux acteurs non européens très bien financés (ex : OneTrust +900M€) ont profité du manque de soutien aux acteurs locaux et de leur difficulté à passer à l'échelle. Ne reproduisons pas les mêmes erreurs et mettons tout en œuvre pour aider nos acteurs à porter le marché des *data spaces*.

L'État peut faciliter l'émergence d'une InfraTech composée entre autres de startups de l'infrastructure des *data spaces* qui s'appuieront sur une vision commune, des standards, des Communs Numériques et une grande interopérabilité. Certaines startups proposent déjà une expertise de grande qualité couvrant de nombreux besoins des *data spaces* (identité / e-Wallets, gestion des modèles légaux et business, intermédiaires de partage de données, gestion des consentements et des contrats pour les partages de données, data pipeline, data management généralistes ou sectoriels, API managers, IA, etc.). Cependant les startups de l'InfraTech éprouvent souvent des difficultés à s'exprimer face aux *hyperscalers*, aux grands groupes ou aux acteurs institutionnels, y compris locaux.

La FrenchTech, et ses licornes, a démontré ces dernières années la capacité de la France à faire émerger des champions dans des applications métier (HealthTech, FinTech, EdTech, etc.), mais elle peine encore à développer les acteurs de l'InfraTech, qui sont pourtant situés à un niveau plus fondamental des chaînes de valeur du numérique, dans tous les secteurs. Si nous la développons, **la multitude d'acteurs InfraTech, européens et français, permettra en retour aux *data spaces* de proposer des projets de confiance, plus souverains.** La France pourra également s'appuyer sur un écosystème fort d'acteurs de l'innovation (Cap Digital, Station F, Paris&Co, etc.) pour accélérer le développement de l'InfraTech et s'assurer de sa diffusion, au travers des acteurs métier de la FrenchTech dans tous les secteurs.

Pour finir, l'État stratège peut consolider l'offre de confiance en soutenant la demande publique. Chacun observera que la puissance considérable acquise par les *hyperscalers* américains s'est significativement nourrie, au cours des dix dernières années, de la commande publique de l'État fédéral des États-Unis d'Amérique, et de ses grandes agences. A titre d'exemple, Amazon web services maintient une position de leader dans les services cloud, avec une part de marché de 32% en mai 2020, position acquise en 2013 lorsque la société a gagné le marché du cloud computing de la CIA pour un montant, record à l'époque, de 600 millions de dollars. Il nous apparaît dès lors indispensable que la commande publique de l'État français, mais également celles des collectivités locales et des opérateurs de réseaux et de services publics, soumis aux mêmes obligations réglementaires, puissent accéder, de manière simple et sans entrave réglementaire excessive, à une offre "numérique de confiance", portée en grande partie par un écosystème de startups, que la plupart des acteurs publics appellent par ailleurs de leurs vœux. Un tel dispositif, qui pourrait être opéré par une agence ou un opérateur public, ouvert aux autres États européens, constituerait un signal fort en direction du marché "Cloud-Data-IA de confiance", et de ses clients publics et privés.

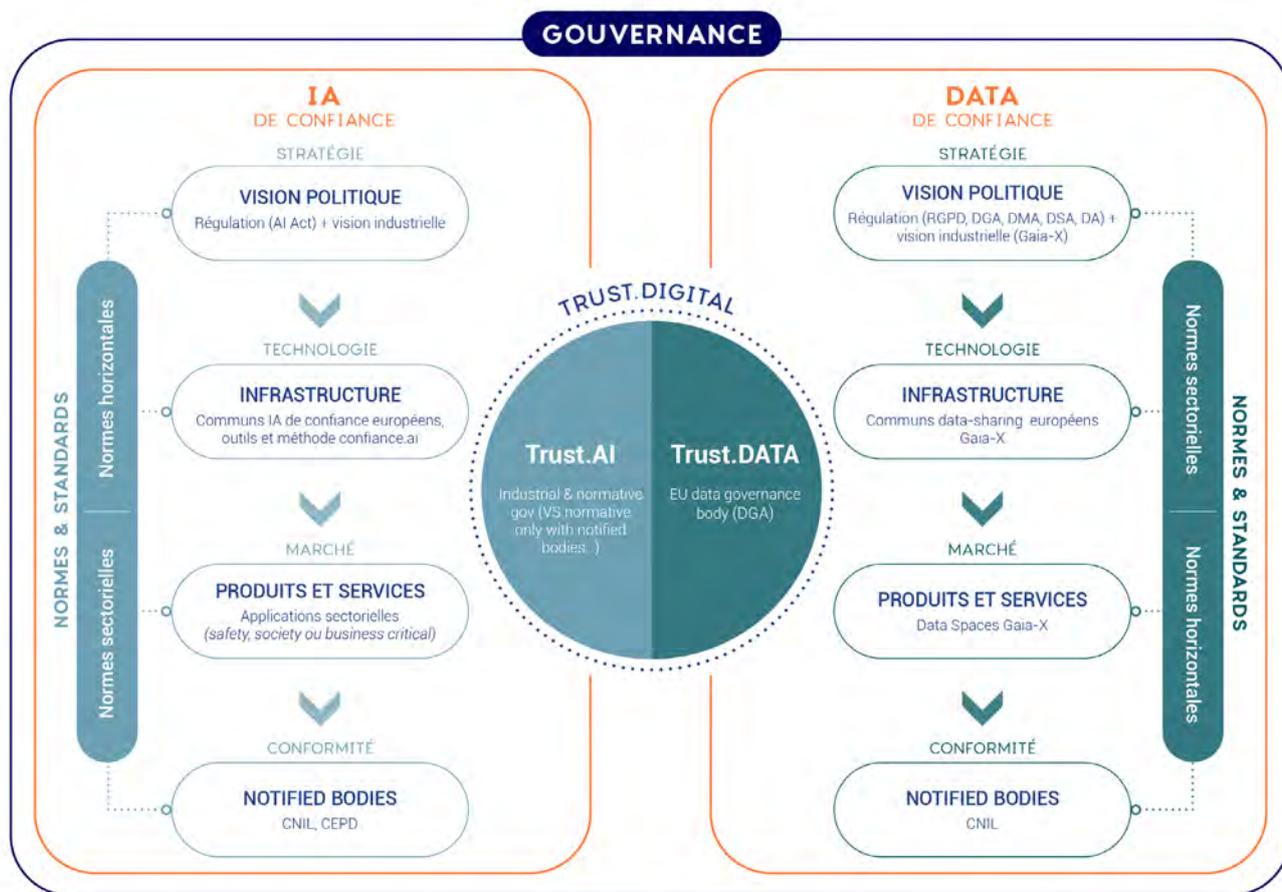
La demande de financement que nous exprimons ici (cf partie droite du schéma ci-dessous) consiste finalement à doter la France des moyens pour peser dans l'opérationnalisation de Gaia-X et ainsi saisir les opportunités économiques que cela fait naître pour nos filières et écosystèmes afférents.



Source: BMWi

Comment ?

- **En finançant une InfraTech** qui permette de :
 - construire une infrastructure commune transversale afin de diminuer le coût de financement des *data spaces* métier
 - ne plus dépendre des solutions étrangères en les désintermédiant
 - enrichir les Communs Numériques
- **En finançant un pilotage commun** qui permette de :
 - soutenir les acteurs pour le développement des normes et labels
 - créer une gouvernance dotée des moyens pour coordonner l'ensemble du programme entre les filières et les États





L'ÉTAT STRATÈGE DOIT
S'ASSURER QUE LES
INFRASTRUCTURES
QUI CONDITIONNENT LES
MARCHÉS D'AVENIR
SOIENT BIEN FINANCÉES
SELON UN PLAN
NATIONAL.

V. PROPOSITIONS POUR UN PROGRAMME NATIONAL

PROGRAMME NATIONAL

C'est à un véritable **Programme « Data de confiance »** auquel nous appelons, et pour lequel nous

FINANCEMENT

Financement d'un Programme Recherche Développement et Innovation (RDI) dédié, supporté par une équipe à temps plein, sur la base des 3 piliers :

PILIER #1 NORMES, CONFORMITÉ ET INTEROPÉRABILITÉ :

Appui aux acteurs pour participer de manière coordonnée au développement des normes et labels de conformité des *data spaces* à l'échelle européenne.

PILIER #2 INFRATECH :

• Communs Numériques :

- identifier les besoins de développement de Communs Numériques de *data spaces* (au-delà des GXFS Gaia-X), soutien des briques déjà développées ou en cours de développement via les AAP déjà en place
- animation de la communauté open source
- coordonner les acteurs infrastructurants et l'écosystème européen (Digital Europe, Simpl)
- mise en œuvre de démonstrateurs

• Acteurs infrastructurants :

- incuber/accélérer les acteurs de l'infrastructure des *data spaces* (startups, grands groupes, etc.)
- dédier un lieu et mettre en place des programmes dématérialisés
- développer l'usage d'intermédiaires de données par les acteurs publics, permettant aux citoyens de contrôler et partager leurs données
- développer une solution d'infrastructure "packagée" s'appuyant sur les Communs Numériques, les multiples solutions des acteurs de l'InfraTech, et les parties prenantes des *data spaces* comme le Hub France Gaia-X, TeraLab⁹³, les pôles de compétitivité (ex : Cap Digital), les PME, les acteurs de la communauté open source (ex : Linagora) et les acteurs de la French Tech (Paris&Co, Lab BPI, Station F, etc.).

PILIER #3 ÉCOSYSTÈMES MÉTIERS :

• Création de *data spaces* métier impliquant des acteurs publics et privés dans les filières jugées prioritaires afin de :

- Promouvoir des **projets phares** (ex : Énergie, Green Deal,...)
- Mettre en place de **bacs à sable réglementaires** pour ces projets
- Développer une couche **IA de confiance** pour les *data spaces*
- **Intégrer des démonstrateurs** d'infrastructure packagée et de Communs Numériques
- **Favoriser des cas d'usage au niveau local** via le tissu des TPE/PME

• Animation et synergies intra et inter *data spaces* :

- Coordination avec les **gouvernances de filières à l'échelle européenne**
- Orchestration avec les **programmes de financement européen** (Digital Europe)
- Participation de la société civile⁹⁴ (concertations, participations à des institutions,...)

Soutien des *data spaces* par la commande publique, via l'incitation dans les appels d'offres à :

- **la cohérence avec Gaia-X** (échelon européen)
- **la pluralité d'acteurs de l'InfraTech** (startups, PME et grands groupes)
- **la réutilisation des Communs Numériques de *data space*** (généralistes et sectoriels).

⁹³ <https://www.teralab-datascience.fr/>

⁹⁴ Par exemple dans des institutions telles que le CNNum ou la CNIL, proposition issue de la note RGPD Acte II : *la maîtrise collective de nos données comme impératif*, Julia Roussoulières et Jean Rérolle, Digital New Deal, 2022

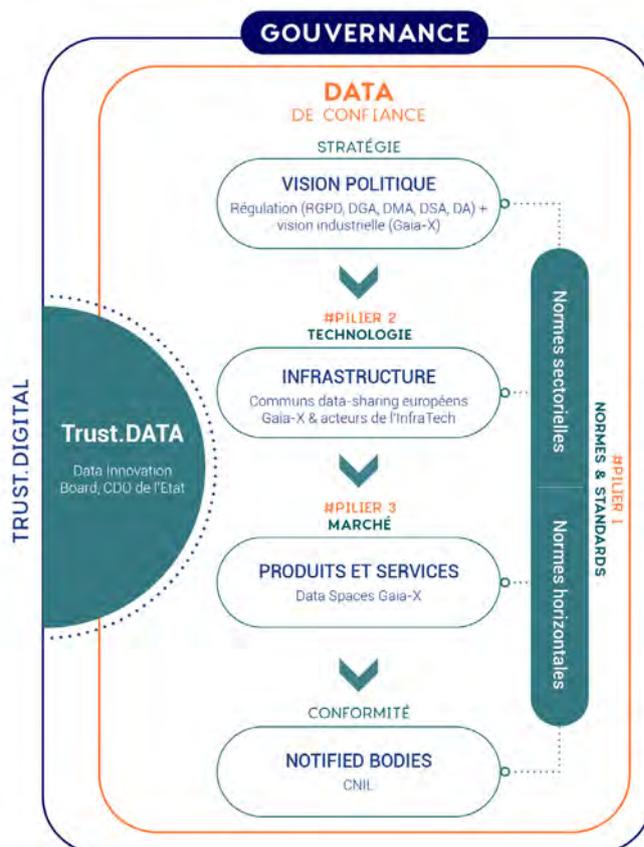
DATA DE CONFIANCE

identifions deux axes prioritaires :

GOVERNANCE

Création d'une gouvernance de la Data de confiance

- **Un organe de gouvernance public** devrait être dédié aux *data spaces*, suivant l'exemple du Data Office espagnol
- **CDO de l'État** : Nous proposons de nous inspirer du modèle espagnol en créant un nouveau rôle de Chief Data Officer de l'État (CDO) chargé de :
 - l'orchestration de l'ensemble de l'écosystème de confiance (Cloud-Data-IA), incluant la question du partage de données et la supervision des projets *data spaces*
 - la coordination public-privé trans-filières au niveau de l'État
 - la représentation de la France dans les instances européennes
- **États Généraux de la donnée**⁹⁵ : Afin de faire comprendre l'opportunité de création de valeur dans toutes les filières, nous proposons d'organiser les États Généraux du partage de données par filière ou par thématique pour :
 - rassembler toutes les catégories d'acteurs (institutionnels, grands groupes, PME/startups) autour de la création de cas d'usage
 - amorcer le développement des *data spaces* (santé, mobilité, données de l'administration, culture, tourisme, éducation, etc.)



⁹⁵ Comme ceux organisés par Digital New Deal et Bpifrance pour la filière tourisme (2 jours avec les acteurs économiques et institutionnels pour aboutir sur une feuille de route opérationnelle : Livres blancs *Partage des données & Tourisme*)

CONCLUSION

Ce livre blanc est le fruit de nombreuses discussions avec les acteurs privés et publics qui participent à la création d'un écosystème numérique français souverain et compétitif. Il tente de traduire le plus pragmatiquement et utilement possible les attentes des acteurs impliqués dans cette grande aventure qui est celle de la donnée, et au-delà de la data, du numérique de confiance dans son ensemble.

Des parties-prenantes parfois inquiètes de notre retard, et par conséquent vigilantes sur les conséquences technologiques, juridiques, et commerciales que cela peut impliquer. Sommes-nous résignés à voir nos données stratégiques, voire d'intérêt général (santé aujourd'hui, éducation ou énergie demain,...) déléguées complètement à des *hyperscalers* américains ou chinois ? Les perturbations géopolitiques actuelles nous invitent pourtant à ne pas choisir le chemin dangereux des dépendances.

Nous sommes en effet à un carrefour de notre histoire, si nous ne saisissons pas maintenant notre chance, nous risquons une nouvelle fois de manquer le train de la révolution digitale. L'Union européenne avec le package réglementaire en cours (DSA, DMA, DGA, DA, IA Act) offre le contexte politique d'émancipation que nous attendions depuis des années, et nous procure même le cadre opérationnel, avec Gaia-X, pour nous arrimer. Nous n'aurons pas d'autres *momentum* aussi favorables avant longtemps...

La France a montré dans son histoire qu'elle pouvait lancer de grands programmes d'infrastructures matérielles (réseau ferré par exemple). Il est temps qu'elle **renoue avec cette ambition en visant la première grande infrastructure immatérielle**. C'est pour cela que nous appelons à un État stratège sur la question centrale de la Data. L'État devrait selon nous orchestrer l'émergence d'une InfraTech du partage de données, basée sur des Communs Numériques, afin de pouvoir accompagner le développement des *data spaces*. Et ce faisant, offrir le **socle technique et politique pour une troisième voie numérique européenne**.

L'Union européenne dispose pour cela d'une opportunité unique de financement, la crise Covid ayant déclenché un **plan européen de relance et résilience de 750 milliards d'euros, dont 20 % sont consacré à la "Boussole numérique"**. Ce "2030 Digital Compass" comprenant une nouvelle structure de gouvernance et « un mécanisme pour organiser avec les États membres les projets multi-pays qui sont nécessaires pour construire la transition numérique de l'Europe dans des domaines critiques ». Il n'est pas trop tard pour **flécher ces milliards vers des programmes existants (Gaia-X) et des programmes manquants (Grand Défi Data)**, pour faire de l'écosystème de confiance le terreau d'une Europe puissance.

CONTRIBUTEURS & REMERCIEMENTS

A l'initiative du Think Tank Digital New Deal, la communauté des acteurs français des *data spaces* (grands groupes, institutionnels, académiques et startups), appelle à un État Stratège sur la question du partage de données de confiance et des *data spaces*.

DIRECTION ET ÉDITORIALISATION

Olivier Dion, *Coordinateur technique Digital New Deal. CEO Onecub (InfraTech), CTO Themis-X (data space tourisme), co-fondateur aNewGovernance et*

Arno Pons, *Délégué Général Digital New Deal, CEO Themis-X (data space tourisme)*

tiennent à remercier pour leur contribution :

CONSEIL ET CONTRIBUTIONS

Henri d'Agrain, *Délégué Général Cigref - Hub France Gaia-X*

Patrick Bezombes, *Président du comité de normalisation IA et Big Data, co-président atelier "Souveraineté numérique" - CEN CENELEC, Vice-président - JTC 21 (IA) et AFNOR*

Yannick Bonhomme, *Responsable valorisation - IRT SystemX, Confiance.ai*

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

David Chassan, *Chief Strategy Officer - 3DS Outscale (InfraTech)*

Antoine Couret, *Président Hub France IA, Président ALEIA (InfraTech) - SaaS collaborative and sovereign AI platform*

Matthias De Bièvre, *Président Visions, Co-fondateur aNewGovernance, Président Prometheus-X - data space éducation et compétences*

Laurent Giovachini, *DGA de Sopra Steria (InfraTech)*

Martine Gouriet, *Directrice des usages numériques - EDF, Data Space Energie*

Pierre Gronlier, *CTO Gaia-X*

Patrick Laurens-Frings, *Directeur des Systèmes d'Information Caisse des Dépôts (data space finance)*

Jennyfer Lecompte, *Directrice des programmes de Recherche, Développement & Innovation CapDigital*

François Marchessaux, *Senior Partner - Franz Partners, expert Data*

Emmanuel Mondon, *Managing Director - Space Coopérative Europe (data space observation de la terre)*

Yves Nicolas, *CTO Sopra Steria*

Michel Paulin, *Directeur général - OVHcloud (InfraTech)*

Sébastien Picardat, *CEO Agdatahub (data space Agriculture)*

Eric Pol, *Président Fondateur aNewGovernance, Bruxelles*

Bruno Sportisse, *PDG INRIA*

Olivier Senot, *Digital Innovation Officer Docaposte (InfraTech), Membre fondateur de Gaia-X*

Marine de Sury, *Directrice de mission Cigref - Hub France Gaia-X*

Anne-Sophie Taillandier, *Institut Mines Télécom - TeraLab - Membre fondateur de Gaia-X. Responsable des services fédérés Gaia-X (GXFS)*

Hubert Tardieu, *Administrateur Indépendant - Gaia-X (ex Président du Conseil d'Administration)*

Sabine Van Haecke-Lepic, *Enseignante Chercheur, Avocate Web3, RegTech*

CONTRIBUTEURS & REMERCIEMENTS

INTERVIEWS ET PARTICIPATIONS COMPLÉMENTAIRES

Delphine Bianchini-Manno, *Consultant Tourism & Hospitality – Themis-X (data space tourisme)*

Hervé Bonazzi, *CEO Archipels (InfraTech) - Digital Trust & Identity, Blockchain, SelfSovereignIdentity*

Claudio Cimelli, *Direction du numérique - Ministère de l'Éducation nationale et de la Jeunesse, coordinateur DASES - Data Space Education & Skills dans GAIA-X*

Jérôme Giusti, *avocat spécialiste en droit de la propriété intellectuelle, en droit du numérique, et en Communs Numériques*

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

Benoît Jaffeux, *Président Fondateur Neobridge (InfraTech) - Editeur logiciel pour la formation et Organisme de Formation*

Sylvain Le Bon, *Co-fondateur Startin'Blox (InfraTech), technologie Solid*

Arnaud Muller, *DG - Cleyrop (InfraTech) - DataHub souverain industriel & modulaire*

Vincent Puren, *Membre de la direction Future4care, cluster multi-disciplinaire de données de santé (data space santé)*

Zokama Sakanga, *Co-fondateur SMART B (InfraTech) - blockchain*

Thomas Saint-Aubin, *CEO Seraphin.legal (InfraTech) - gestion des contrats, Legaltech*

Matthieu Sonnati, *Président Inokufu (InfraTech), cofondateur - Prometheus-X (data space éducation & compétences)*

Alexandre Zapolsky, *Président-fondateur – Linagora*

OLIVIER DION

Ingénieur diplômé de Telecom SudParis / Hong Kong University of Science and Technology en 2007, passionné d'intelligence artificielle et précurseur de l'Open Data en France, Olivier Dion a fondé en 2011 la startup Onecub, pionnière dans le domaine de la portabilité et du partage des données personnelles. Il a également co-fondé aNewGovernance, acteur clé des data spaces, avec de multiples parties-prenantes de la communauté européenne des données. Il a été impliqué dans la construction du RGPD avec la CNIL (droit à la portabilité), et dans la stratégie des données européennes avec la Commission européenne. Il pilote aujourd'hui le data space tourisme Themis.



ARNO PONS

Délégué général du « Think-tank Digital New Deal », co-auteur de quatre notes sur le numérique de confiance (Cloud de confiance, Infrastructures du numérique de confiance, IA de confiance, et Data de confiance). Il a fondé en 2021 l'activité « Do Tank - Digital New Deal », dédiée aux enjeux de coopération pour accompagner les entreprises et collectivités à structurer les filières en écosystèmes digitaux via des alliances technologiques. Il lance la première initiative de place du Do Tank en co-crédant en 2022 Themis (*data space* tourisme comprenant soixante entités partenaires). Il avait créé auparavant plusieurs start-up (Checkfood-gaspillage alimentaire, Medicimo au Canada,...), et également enseigné à SciencesPo sur les enjeux de souveraineté numérique liés à la centralisation des pouvoirs par les *Big Tech*.

DIGITAL NEW DEAL

LE THINK-TANK DE LA NOUVELLE DONNE

Digital New Deal accompagne les décideurs privés et publics dans la création d'un Internet des Lumières, Européen et Humaniste. Notre conviction est que nous pouvons offrir une 3^{ème} voie numérique en visant un double objectif : défendre nos valeurs en proposant une nouvelle régulation contre la centralisation des pouvoirs ; et défendre nos intérêts en créant les conditions de la coopération face à la captation de la valeur par les « Big Tech ».

Notre activité de publication a pour vocation d'éclairer de manière la plus complète possible les évolutions à l'œuvre au sein de enjeux de « souveraineté numérique », dans l'acception la plus large du terme, et d'élaborer des pistes d'actions concrètes, voire opérantes via le Do tank, à destination des organisations économiques et politiques.

LE CONSEIL D'ADMINISTRATION

Olivier Sichel (président fondateur) et Arno Pons (délégué général), pilotent les orientations stratégiques du think-tank sous le contrôle régulier du conseil d'administration.

Forts de leur intérêt commun pour les questions numériques, les membres du Conseil d'administration ont décidé d'approfondir leurs débats en formalisant un cadre de production et de publication au sein duquel la complémentarité de leurs expériences pourra être mise au service du débat public et politique. Ils s'impliquent personnellement dans la vie de Digital New Deal, notamment dans le choix des rapports et de leurs rédacteurs. Il sont les garants de notre indépendance, académique et économique.



SÉBASTIEN BAZIN
PDG AccorHotels



NATHALIE COLLIN
DG branche Grand Public et
Numérique Groupe La Poste



NICOLAS DUFOURCQ
DG de Bpifrance



AXELLE LEMAIRE
Ex-Secrétaire d'Etat
du Numérique et de
l'Innovation



ALAIN MINC
Président AM Conseil



DENIS OLIVENNES
DG Libération



YVES POILANE
DG Ionis Education Group



ARNO PONS
Délégué général du think
tank Digital New Deal



JUDITH ROCHFELD
Professeure agrégée de Droit,
Panthéon Sorbonne



OLIVIER SICHEL
Président Digital New Deal
DGA Caisse des Dépôts



BRUNO SPORTISSE
PDG Inria



ROBERT ZARADER
PDG Bona fidé

IA de confiance, opportunité stratégique pour une souveraineté industrielle et numérique | Julien Chiaroni, Arno Pons - *juin 2022*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

La santé face au tsunami des NBIC et aux plateformes | Laurent Alexandre - *juin 2016*

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

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



THINK-TANK
DIGITAL
NEW DEAL

septembre 2022

www.thedigitalnewdeal.org