

Vers une infrastructure européenne de confiance pour le partage souverain des données de recherche

Synthèse exécutive

L'Europe produit des volumes considérables de données de recherche à haute valeur stratégique dans des domaines comme la santé, l'industrie, le droit, les politiques publiques ou l'intelligence artificielle. Pourtant, malgré les investissements engagés dans les infrastructures numériques et dans l'European Open Science Cloud (EOSC), **le partage transfrontalier de ces données demeure fragmenté, juridiquement complexe et difficile à industrialiser.**

Ce paradoxe constitue aujourd'hui un frein majeur à la souveraineté scientifique et numérique européenne.

Le problème n'est pas l'absence de données, ni même l'absence de cadre réglementaire. **L'Europe dispose d'un environnement juridique robuste (RGPD, Data Governance Act, Data Act, AI Act) ainsi que d'infrastructures de recherche performantes.** Le véritable blocage réside dans la difficulté à transformer ces règles en mécanismes opérationnels de confiance.

Aujourd'hui, la gouvernance du partage des données repose encore largement sur des instruments documentaires : **contrats de partage, licences, clauses RGPD, autorisations éthiques, politiques institutionnelles.** Ces outils sont juridiquement indispensables mais restent essentiellement statiques, **conçus pour une lecture humaine, difficilement interopérables et quasi impossibles à exécuter automatiquement.**

Cette situation crée plusieurs obstacles structurels.

D'abord, un déficit de souveraineté opérationnelle. Une fois l'accès accordé, les producteurs de données — universités, laboratoires, institutions publiques ou consortiums — perdent souvent la maîtrise effective des conditions d'usage de leurs datasets. Cela est particulièrement critique pour les données sensibles : santé, données publiques, données industrielles, corpus juridiques ou datasets d'entraînement IA. **Les infrastructures actuelles permettent l'accès, mais rarement le contrôle continu des usages, des restrictions de réutilisation ou des interdictions d'entraînement IA.**

Ensuite, une forte fragmentation juridique transfrontalière. Les collaborations européennes doivent naviguer entre RGPD, règles nationales, contraintes sectorielles, politiques propres aux établissements et exigences éthiques. **Le problème n'est pas juridique, mais opérationnel : comment rendre cette complexité gouvernable à l'échelle ?**

À cela s'ajoute un **déficit de traçabilité**. Les chercheurs doivent pouvoir vérifier l'origine des données, leur historique, leurs transformations et les conditions de réutilisation successives. **Sans provenance fiable, la reproductibilité scientifique, la conformité réglementaire et la confiance dans l'IA deviennent fragiles.**

Enfin, **EOSC** lui-même souffre encore d'un **déficit d'interopérabilité** avec les nouveaux mécanismes de gouvernance de confiance développés dans les autres Data Spaces européens. Or l'avenir du partage des données repose précisément sur cette convergence.

Face à ce constat, **notre consortium propose une infrastructure européenne de gouvernance programmable permettant de transformer la conformité documentaire en gouvernance opérationnelle exécutable.**

Notre approche repose sur une architecture en quatre couches.

1 **La première couche est celle du commun technique**, fondé sur le concept de **Data Contract**. Il s'agit de transformer les accords juridiques statiques en **politiques machine-readable et exécutables**. Ce socle repose sur quatre briques technologiques complémentaires : **ODRL** pour exprimer les **droits, obligations et interdictions** ; **JSON-LD** pour la provenance et la **traçabilité** ; la modélisation des finalités RGPD afin d'intégrer les contraintes juridiques directement dans les politiques ; et enfin la **gouvernance des chaînes de services** pour garantir que les règles suivent les traitements dans des architectures distribuées.

Ce commun technologique est neutre, interopérable et réutilisable dans toutes les disciplines scientifiques.

2 **La deuxième couche est celle de la fédération de nœuds souverains**. Notre projet ne repose pas sur une centralisation des données, mais sur une **architecture distribuée dans laquelle chaque producteur conserve la maîtrise de ses actifs**. Les données restent chez leur propriétaire ; **seules circulent** les politiques d'usage, les **métadonnées** de gouvernance et les attestations de conformité. Le pilote initial s'appuie sur l'**écosystème juridique déjà structuré autour du Legal Data Space et des travaux AI is Law**, avant d'être étendu à plusieurs disciplines afin de démontrer le caractère générique du standard.

Cette approche répond directement à l'**exigence centrale de l'appel EOSC : garantir un contrôle effectif des data owners.**

3 **La troisième couche est celle de la confiance certifiée**. L'interopérabilité technique ne suffit pas à créer un réseau de confiance. C'est pourquoi nous proposons un mécanisme d'audit indépendant, porté avec des partenaires spécialisés comme Data Craft / Labelia, permettant d'évaluer les nœuds participants sur leurs capacités de **gouvernance**, d'**auditabilité**, de **souveraineté** et de **conformité opérationnelle**. Les nœuds conformes se verront attribuer un

label de confiance, transformant une simple fédération technique en un véritable réseau trusted.

Cette couche répond explicitement à l'exigence de "trusted framework" mise en avant par l'appel.

4 Enfin, la quatrième couche est celle du **support EOSC**, conçue comme un centre opérationnel léger facilitant l'adoption du framework. Il ne s'agit pas de créer une nouvelle bureaucratie, mais de mettre à disposition un helpdesk, un registre de gouvernance, une bibliothèque de templates de Data Contracts et des mécanismes d'intégration progressive avec EOSC Federation.

L'ambition est double.

À court terme, démontrer qu'il est possible d'automatiser la gouvernance du partage de données sensibles dans un cadre souverain, traçable et interopérable.

À moyen terme, proposer à EOSC une couche de gouvernance de confiance réutilisable par l'ensemble de l'écosystème européen de recherche.

Cette approche apporte plusieurs bénéfices stratégiques : souveraineté by design, conformité automatisable, réduction de la complexité juridique, traçabilité native, compatibilité avec les usages IA, et interopérabilité avec les autres Data Spaces européens.

L'Europe n'a pas uniquement besoin de nouvelles infrastructures de stockage ou de calcul.

Elle a besoin d'une infrastructure européenne de confiance pour gouverner la circulation des données sensibles.

Notre proposition vise précisément à construire cette couche manquante.

Du PDF au policy-as-code. De la conformité déclarative à la gouvernance exécutable.
Du partage fragmenté à une souveraineté scientifique européenne de confiance.