

Le rôle du Gouvernement en tant que fournisseur de données pour l'Intelligence Artificielle

**Phase 2 : Un outil pratique pour le partage
de données gouvernementales pour l'IA**

Novembre 2024



GPAI

THE GLOBAL PARTNERSHIP
ON ARTIFICIAL INTELLIGENCE

Ce rapport a été élaboré avant l'intégration du Partenariat mondial sur l'intelligence artificielle (PMAI) et de l'Organisation de coopération et de développement économiques (OCDE) à la mi-2024. Par conséquent, il n'a pas été soumis à l'approbation des membres du PMIA et de l'OCDE et ne doit pas être considéré comme reflétant leurs positions.

Remerciements

Ce rapport a été élaboré dans le cadre du projet «Le Rôle du Gouvernement en tant que Fournisseur de Données pour l'Intelligence Artificielle», sous la direction des Co-responsables du Projet et les conseils du Groupe Consultatif du Projet, avec le soutien du Groupe de travail d'experts sur la gouvernance des données du GPAI. Le Groupe de Travail d'Experts sur la Gouvernance des Données du GPAI a accepté de déclassifier ce rapport et de le rendre public.

Co-Responsables :

Ching-Yi Liu*, Université Nationale de Taiwan

Jhalak Kakkar*, Centre pour la Gouvernance des Communications, en congé depuis juin 2024

Avik Sarkar*, Ecole Indienne de Commerce

Ce rapport a été rédigé par : **Annys Rogerson†**, **Sulamaan Rahim‡**, **Pablo Fuentes Nettel‡**, **Richard Stirling‡**, and **Yaseen Ladak‡** d'Oxford Insights; **Muchiri Nyaggah†**, Local Development Research Institute; **Natalia Carfi†**, Open Data Charter; **Juan Carlos Lara†**, Derechos Digitales.

Nous tenons à remercier les membres du groupe consultatif pour leur soutien actif tout au long du projet **Thierry Warin*** (HEC Montréal), **Seong Oun Hwang*** (Gachon University), **Andrew Sporle*** (iNZight Analytics Ltd), **Zümrüt Müftüoğlu*** (Yildiz Technical University), **Jaco Du Toit**** (UNESCO), **Kudakwashe Dandajena*** (AIMS-II), **Zee Kin Yeong*** (IMDA), **Kim McGrail*** (University of British Columbia), **Mikael Jensen*** (D-Seal), and **Ulises Cortes*** (Universitat Politècnica de Catalunya).

Nous sommes également reconnaissants à des juristes de l'AICoE de Taiwan pour leur examen du texte, **Chuan-Feng Wu†**, **Tai-Jan Huang†**, **Wen-Tsong Chiou†**, **Ching-Fu Lin†** and **Ting-Chi Liu†**.

Nous tenons à remercier **Lina Jahidi‡** d'Oxford Insights pour la traduction de ce rapport en français.

GPAI tient à saluer les efforts inlassables de ses collègues du Centre International d'Expertise de Montréal en Intelligence Artificielle (CEIMIA) et du Groupe de travail d'experts sur la gouvernance des données du GPAI. Nous sommes particulièrement reconnaissants du soutien d'**Antoine Glory**, **Nathalie Noël**, **Stephanie King**, et **Stefan Janusz** du CEIMIA, ainsi que du dévouement des Coprésidents du Groupe de travail, **Shameek Kundu** et **Bertrand Monthubert**.

* Expert

** Observateur

† Spécialiste invité

‡ Parties contractantes des Centres d'Excellence (Cde) pour contribuer aux projets

Citation

GPAI 2024. Le rôle du Gouvernement en tant que fournisseur de données pour l'Intelligence Artificielle : Feuille de route de mise en œuvre, Novembre 2024, Partenariat Mondial sur l'IA.

Table des matières

Objectif de la feuille de route.....	1
Vue d'ensemble de la feuille de route.....	3
Un cadre pour le partage de données pour l'IA.....	4
Utilisation de la feuille de route.....	8
Évaluation initiale.....	9
Liste de contrôle pour le partage de données gouvernementales pour l'IA.....	13
Étape de décision : Cas d'usage.....	16
Étape de décision : Mécanisme.....	26
Étape de décision : Structure de gouvernance.....	36
Études de cas pilotes.....	44
Annexe 1 : Ressources supplémentaires.....	59
Annexe 2 : Méthodologie complète.....	83
Bibliographie.....	86

Objectif de la feuille de route

Cette feuille de route vise à aider les équipes gouvernementales à décider quand et comment partager des données pour l'innovation en intelligence artificielle (IA). Pour ce faire, elle fournit une liste de contrôle, accompagnée de conseils pratiques et d'exemples, destinée aux équipes gouvernementales travaillant sur des initiatives de partage de données pour l'IA. Cette liste de contrôle a pour but d'orienter ces initiatives vers la création de bénéfices publics, tout en respectant les droits des individus et des communautés concernés, tout en contribuant aux objectifs d'égalité et de justice des données.

La feuille de route s'adresse aux équipes gouvernementales de tout niveau qui s'intéressent ou sont déjà engagées dans des projets de partage de données avec des développeurs d'IA externes. Elle peut être utilisée par les équipes pour :

- Élaborer un plan pour une nouvelle initiative de partage de données pour l'IA dès le début.
- Développer un plan pour l'aspect partage de données de projets existants et plus vastes déjà en cours au sein de l'équipe.
- Évaluer systématiquement les initiatives de partage de données existantes.

Motivations du partage de données pour le développement de l'IA

Les gouvernements détiennent des données précieuses pour le développement de l'IA. Cela inclut des données sur la société, l'économie et des données administratives, générées par la prestation de services et les exercices de statistiques nationales. Ces données pourraient être utilisées par des entreprises, des organismes de recherche et des organisations à but non lucratif pour le développement de recherches, de produits et de services basés sur l'IA, à utiliser à l'intérieur et à l'extérieur du gouvernement.

Reconnaissant ce potentiel, de plus en plus d'équipes gouvernementales commencent à explorer comment partager des données de manière responsable et efficace avec les développeurs d'IA. Les équipes ont des ambitions très diverses en matière de partage de données avec ces développeurs, notamment :

- Faciliter les applications d'IA pour améliorer la prestation de services gouvernementaux (Berryhill et al, 2019).
- Commander ou permettre la recherche axée sur l'IA pour mieux comprendre des problèmes sociaux complexes et éclairer les décisions politiques (Ziesche, 2023).
- Soutenir l'innovation des produits et services d'IA en dehors du gouvernement afin de développer le secteur national de l'IA et soutenir son adoption (Ceulemans et al, 2021).
- Cibler les problèmes dans des secteurs clés et des enjeux de développement que les produits ou services d'IA peuvent aider à résoudre (GPAI, 2022a).

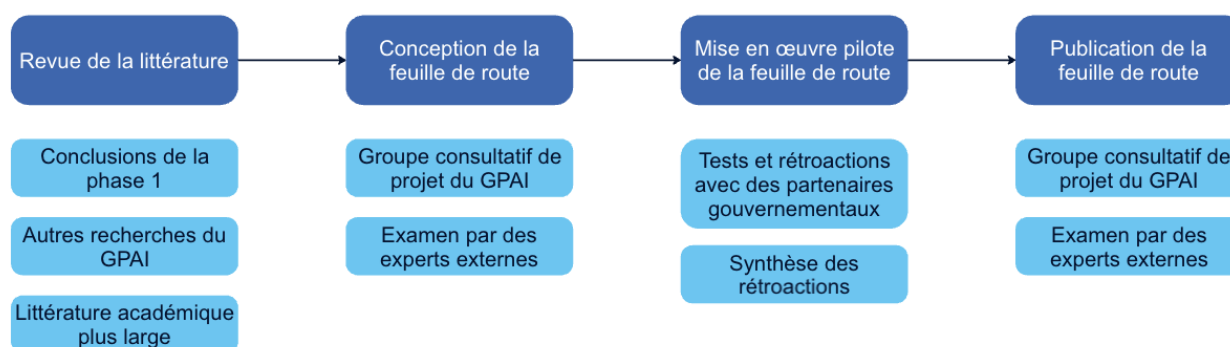
La première phase du projet du GPAI "Le gouvernement en tant que fournisseur de données pour l'IA" a documenté la manière dont plusieurs gouvernements partagent déjà des données pour l'IA (GPAI, 2023a). Le gouvernement de Taïwan facilite le partage de données de santé entre les citoyens et des développeurs d'applications d'IA tiers approuvés via son *Health Passbook* pour permettre le développement de technologies de santé basées sur l'IA. Le gouvernement de la Colombie soutient le partage de données gouvernementales et de données d'agriculteurs sur sa plateforme Aclímate afin de permettre la fourniture d'informations basées sur l'IA sur le climat et les conditions environnementales aux agriculteurs. Des exemples d'initiatives de partage de données pour l'IA sont également donnés tout au long du rapport.

Malgré ces ambitions, de nombreuses équipes sont incertaines en ce qui concerne quand et comment les données peuvent être partagées avec des parties externes. Cet exercice doit être mené conformément aux principes des droits de la personne, de la vie privée et de la protection des données, de l'inclusion, de la diversité, de l'innovation et de la croissance économique. Pour ce faire, les équipes sont souvent confrontées à un certain nombre de défis, notamment l'établissement de la confiance du public, la certitude réglementaire et le renforcement des capacités techniques (GPAI, 2023). Cette feuille de route vise à aider les équipes gouvernementales à naviguer dans leurs initiatives de partage de données et à relever ces défis.

Élaboration de la feuille de route

Oxford Insights a été mandatée pour élaborer une feuille de route pratique destinée aux gouvernements, visant à faciliter le partage de données pour l'innovation en IA, et pour tester cette feuille de route avec des partenaires pilotes.

La feuille de route intègre les conclusions de la première phase du projet du GPAI, “Le Gouvernement en tant que fournisseur de données pour l'IA”. Cette première phase a utilisé quatre études de cas de partage de données gouvernementales pour l'IA afin de développer un ensemble de principes recommandés aux les gouvernements (GPAI, 2023a). Un premier jet de la feuille de route a été élaboré, complété par les apports d'une revue de littérature plus large et les commentaires d'experts mondiaux en données ouvertes, en partage de données et en IA.



Afin de tester l'utilisabilité de la feuille de route, celle-ci a ensuite été testée en pilote auprès d'équipes gouvernementales dans trois pays, à différents niveaux de gouvernement :

- L'Agence pour le Gouvernement Électronique et l'Information de l'Uruguay.
- Le Bureau de la Transformation Numérique de la Présidence de la Turquie.
- Le Gouvernement de l'État de Jigawa.

Nous avons sélectionné ces partenaires pilotes après avoir lancé un appel ouvert à manifestation d'intérêt auprès d'équipes gouvernementales du monde entier. Notre objectif était de nous assurer que la feuille de route est applicable dans différents contextes. Par conséquent, nous avons sélectionné les partenaires pilotes de manière à ce qu'ils représentent, entre autres critères, un éventail diversifié de zones géographiques, de niveaux d'expérience en matière de partage de données et de secteurs d'intérêt pour les projets pilotes.

Cette version finale a été itérée pour refléter les précieuses perspectives et les leçons tirées de ces pilotes, garantissant qu'elle répond aux objectifs et aux besoins divers des initiatives gouvernementales de partage de données pour l'IA.



Vue d'ensemble de la feuille de route

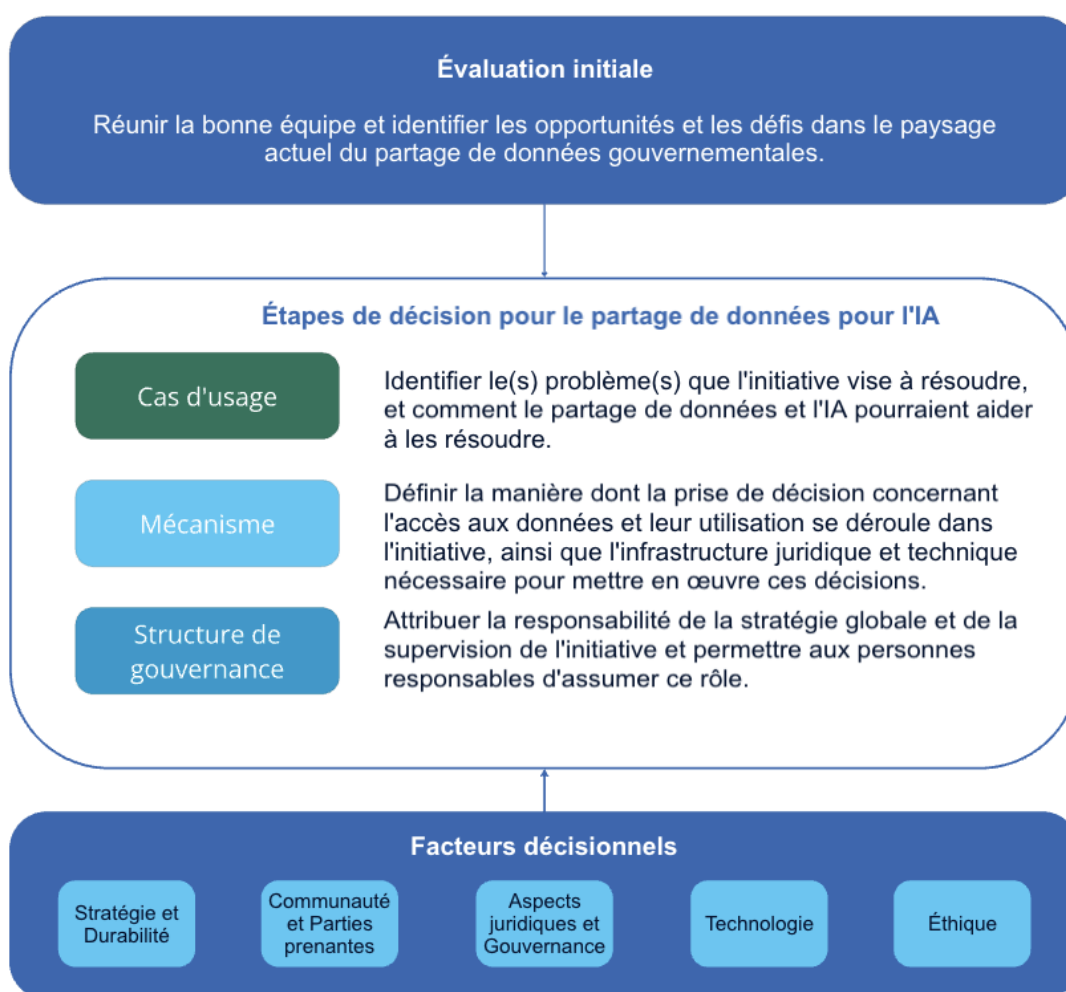
Section du rapport	Description de la section
Cadre pour le partage de données pour l'IA	Présente aux utilisateurs le cadre de prise de décision utilisé dans la feuille de route
Utilisation de la feuille de route	Conseille les utilisateurs sur la manière d'utiliser la feuille de route pour faire avancer leurs projets de partage de données pour l'IA.
Évaluation initiale	Fournit à l'utilisateur les premières étapes recommandées pour son projet, consistant à rassembler la bonne équipe et à évaluer le paysage actuel du partage de données gouvernementales.
Liste de contrôle de la feuille de route	Présente un ensemble de questions que les utilisateurs doivent se poser pour progresser à chaque étape du cadre de prise de décision.
Guide de la liste de contrôle	Explique pourquoi chaque question de la liste de contrôle de la feuille de route est incluse et décrit ce qui peut être nécessaire pour y répondre. Le guide comprend également des exemples internationaux, des activités recommandées et des enseignements tirés des partenariats pilotes afin de fournir aux utilisateurs les ressources nécessaires pour aborder la question.
Études de cas pilotes	Présente les enseignements tirés des partenariats pilotes ainsi qu'une étude de cas des parcours de mise en œuvre de la feuille de route et des résultats pour chacun des trois partenaires pilotes.
Ressources supplémentaires	Fournit des ressources supplémentaires, telles que des documents d'atelier et des explications de termes clés, qui ont été utilisés lors des partenariats pilotes et pourraient aider les utilisateurs à répondre aux questions de la liste de contrôle.
Méthodologie complète	Décrit la méthodologie détaillée du développement de la feuille de route, ainsi que les changements clés apportés à la suite des enseignements tirés des partenariats pilotes.

Un cadre pour le partage de données pour l'IA

La feuille de route accompagne les équipes gouvernementales dans le développement d'une initiative de partage de données pour l'IA, en les guidant d'abord à travers une évaluation initiale, puis à travers trois étapes de décision pour leurs projets.

Pour chaque étape de décision, la feuille de route les divise en considérations clés pour développer des initiatives efficaces et responsables. Ces considérations sont regroupées en cinq facteurs décisionnels et sont présentées sous forme de questions dans une liste de contrôle que les équipes doivent compléter.

Cadre de la feuille de route pour le partage de données gouvernementales pour l'IA



Évaluation initiale

La feuille de route commence par une évaluation initiale de l'état de préparation pour entreprendre une initiative de partage de données pour l'IA. Le but de cette évaluation est, d'une part, de s'assurer que l'équipe et les parties prenantes clés ont la capacité appropriée pour entreprendre l'initiative, et d'autre part, d'identifier les facilitateurs et les défis pour l'initiative dans le paysage du partage de données du gouvernement.

Cas d'usages

Les cas d'usages relient l'initiative aux problèmes qu'elle vise à résoudre. Sans des cas d'usage clairs, pertinents et réalisables, l'initiative n'aura pas d'impacts concrets.

Dans cette feuille de route, un cas d'usage pour une initiative de partage de données pour l'IA spécifie le problème que l'équipe souhaite résoudre, les applications d'IA qui pourraient aider à le résoudre, les données qui doivent être partagées pour faciliter ces applications, et qui doivent être les utilisateurs des données.

Les objectifs des questions sur les cas d'usage dans la liste de contrôle sont de :

- (1) aider l'équipe à identifier les cas d'usage potentiels;
- (2) établir les limites légales et éthiques dans lesquelles les données peuvent être partagées et à quelles fins;
- (3) prioriser les cas d'usage avec lesquels aller de l'avant.

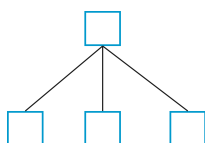
Mécanismes de partage de données

Les mécanismes de partage de données définissent comment la prise de décision concernant l'accès aux données et leur utilisation s'effectue dans l'initiative, ainsi que l'infrastructure juridique et technique nécessaire pour mettre en œuvre ces décisions. Un des facteurs principaux qui influence la conception des mécanismes est la création de la confiance parmi les participants à l'initiative, confiance nécessaire à leur participation.

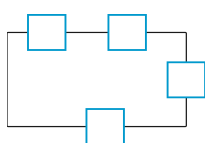
Plusieurs mécanismes émergent pour permettre aux gouvernements de fournir des données pour l'IA. Il existe de nombreuses façons de catégoriser ces mécanismes. Dans cette feuille de route, ils sont décomposés en trois types de mécanismes, basés sur la manière dont les données circulent entre les participants à l'initiative.



Partage de données individuel : accords entre deux parties individuelles, qui sont les utilisateurs et/ou les contrôleurs des données. Par exemple : un contrat privé ou un partenariat public-privé.



Partage de données un-à-plusieurs : accords entre une seule partie, qui est le contrôleur des données ou représente les contrôleurs, et de nombreux utilisateurs de données. Par exemple : une initiative de données ouvertes, un bac à sable (sandbox) ou une initiative¹ de gestion de données (data stewardship).



Partage de données plusieurs-à-plusieurs : accords entre de nombreuses parties qui sont des contrôleurs de données et de nombreuses parties qui sont des utilisateurs de données. Il peut s'agir des mêmes parties. Par exemple : un bien commun de données (data commons) ou un marché de données (data marketplace).

¹ Chacune de ces catégories est explorée plus en détail dans la [Taxonomie des Mécanismes de Partage de Données](#), disponible dans l'annexe des Ressources Complémentaires.

Participants aux initiatives de partage de données pour l'IA

Cette feuille de route fait référence aux participants de l'initiative en tant qu'acteurs suivants (GPAI, 2022b) :

- **Les contributeurs de données** sont les personnes physiques ou morales qui sont les sujets de l'information (personnes concernées), dont l'activité engendre la création de nouvelles données, ceux qui rassemblent des données, par exemple dans un jeu de données, et ceux qui contribuent de manière plus indirecte, par exemple en fournissant un appareil qui génère des données.
- **Les contrôleurs de données** sont les acteurs qui détiennent les données et, selon le mécanisme, ils peuvent également décider des finalités et des moyens de leur traitement. Dans ce cas, ils agissent aussi comme **gestionnaires de données**. Il peut y avoir des co-contrôleurs qui exercent conjointement les décisions relatives aux données.
- **Les sous-traitants de données** sont des prestataires de services qui traitent des données pour le compte d'un contrôleur de données.
- **Les utilisateurs de données** sont les acteurs qui reçoivent des données des contrôleurs de données et/ou des gestionnaires de données.
- **Les intermédiaires de données** sont des fournisseurs de services de partage de données, tels que les mandataires de données (data trustees) ou les fournisseurs de marchés de données (data marketplaces).
- **Les bénéficiaires de services** sont les acteurs qui utilisent des services basés sur des données.
- **D'autres acteurs pertinents** sont les régulateurs et les décideurs politiques qui définissent les cadres politiques, législatifs et réglementaires.

L'objectif des questions sur les mécanismes dans la liste de contrôle est de clarifier les rôles des participants et d'évaluer l'infrastructure de données et l'infrastructure juridique nécessaires. Cela permet d'adapter le mécanisme à l'initiative et au contexte gouvernemental.

Structures de gouvernance

Les structures de gouvernance définissent qui sera responsable de la stratégie globale et de la supervision de l'initiative, et comment cette responsabilité sera exercée. Un des facteurs principaux dans la conception d'une structure de gouvernance est le maintien de la confiance des participants et du public une fois l'initiative opérationnelle.

L'objectif des questions sur la structure de gouvernance dans la liste de contrôle est d'attribuer les responsabilités en matière de stratégie et de supervision, et de s'assurer que les processus et les mécanismes sont en place pour que les responsables puissent exercer ce rôle.

Facteurs décisionnels

Chacune des trois étapes de décision est décomposée en considérations relatives à cinq facteurs de décision qui s'appuient sur les recommandations de la première phase de ce projet et des projets plus larges du groupe de travail du GPAI sur la gouvernance des données (GPAI, 2020 ; GPAI, 2022a ; GPAI, 2022b ; GPAI, 2023a ; GPAI, 2023b).

Stratégie et Durabilité

Les cas d'usages ciblent des problèmes qui constituent une priorité stratégique ou de développement pour le gouvernement, et dont la résolution apporterait des bénéfices publics significatifs. L'initiative dispose d'un financement durable et le modèle financier favorise une distribution équitable des bénéfices.

Communauté et Parties Prenantes

Les besoins et les perspectives des participants à l'initiative, des personnes concernées par les données et de tout autre acteur affecté sont pris en compte dès le début du développement et se reflètent dans une conception qui répond aux exigences d'égalité et de justice² des données. La transparence est maintenue pour tous ces acteurs tout au long de la durée de vie de l'initiative.

Aspects juridiques et Gouvernance

L'initiative respecte les cadres législatifs et réglementaires, qui intègrent les droits de l'homme et les droits liés aux données des personnes concernées, et sont mis en œuvre par des institutions compétentes. Les acteurs publics et privés pertinents ont des obligations claires et il existe des mécanismes qui permettent d'assurer leur redevabilité.

Technologie

Les données sont partagées à l'aide d'une infrastructure sûre et sécurisée qui est continuellement surveillée. Les données répondent aux exigences de qualité et d'accessibilité de leurs utilisateurs. Les risques de sûreté et de sécurité sont identifiés et atténués par des analyses d'impact, des audits et des techniques de gouvernance technique. Il existe des processus justes et robustes pour approuver ou acquiescer des utilisateurs de données.

Éthique

L'initiative n'expose pas les individus et les communautés à des risques de préjudice, d'exclusion ou de biais, que ce soit par le partage de données ou par la prise de décision algorithmique. Des analyses d'impact, la supervision humaine, une conception centrée sur l'utilisateur et d'autres processus intégrés sont utilisés pour s'assurer que les systèmes d'IA ne causent pas de préjudice mais sont utilisés de manière à promouvoir l'équité et à apporter des bénéfices publics.

² Le rapport 2022 du GPAI sur la justice des données demande aux décideurs politiques de reconnaître et de réagir aux répartitions inégales des opportunités et des préjudices qui suivent les schémas historiques d'inégalités sociales et géographiques, tant à l'intérieur qu'entre les pays, en raison de l'extraction de données et des systèmes à forte intensité de données (GPAI, 2022c).

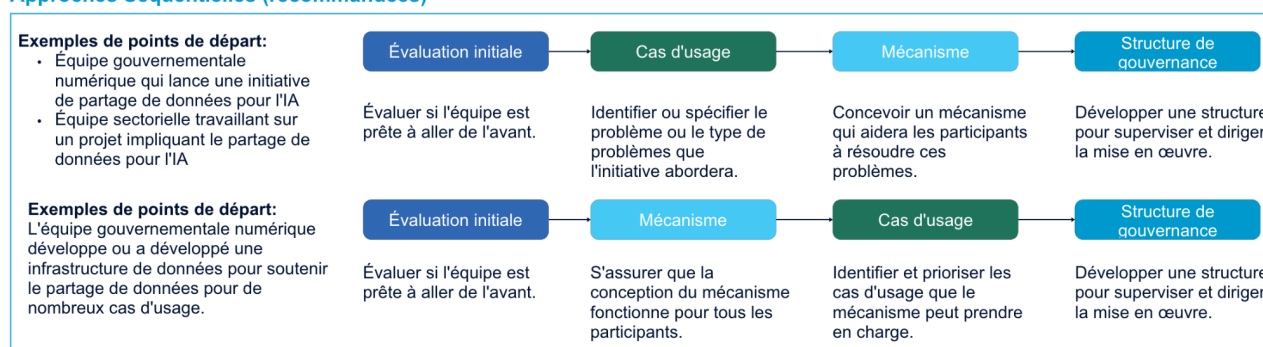
Utilisation de la feuille de route

Parcours des utilisateurs

Les équipes gouvernementales se trouvent à différentes étapes de leur parcours de partage de données : certaines ont déjà identifié des cas d'usage pour le partage de données en IA ou développé une infrastructure de partage de données réutilisable, tandis que d'autres ne font que commencer. La feuille de route est conçue pour être utilisée par des personnes à tous les stades de ce parcours. Pour répondre à ces besoins variés, la feuille de route est donc conçue pour être utilisée de différentes manières. En fonction de leurs objectifs et de leurs points de départ, les équipes peuvent adopter :

- **Une approche séquentielle** : Les équipes peuvent suivre la feuille de route de manière linéaire, depuis le tout début ou en partant d'un cas d'usage ou d'un mécanisme déjà identifié. Nous recommandons aux équipes de compléter l'évaluation initiale, puis de passer aux questions de la liste de contrôle. Selon le point de départ de l'équipe, celle-ci peut commencer par les sections sur le cas d'usage ou sur le mécanisme.

Approches Séquentielles (recommandées)



- **Une approche de revue de projet** : Les équipes peuvent passer en revue des projets de partage de données existants ou améliorer leurs pratiques et leurs orientations en matière de partage de données pour l'IA de manière plus générale.
- **Une approche de résolution de problèmes** : Les équipes peuvent aborder des défis spécifiques au sein d'un projet de partage de données pour l'IA en utilisant les conseils inclus dans la liste de contrôle.



Évaluation initiale

L'objectif de l'évaluation initiale est de comprendre l'étendue des capacités de l'équipe et de l'institution pour entreprendre une initiative de partage de données pour l'IA. Il vise à rassembler la bonne équipe, à soutenir la formation de coalitions avec d'autres parties prenantes pertinentes et à identifier les facilitateurs et les défis pertinents pour l'initiative dans le paysage gouvernemental du partage de données.

La section sur l'évaluation initiale fournit des conseils sur les exigences institutionnelles pour aller de l'avant avec l'initiative, la structure et les compétences de l'équipe, ainsi que les ressources nécessaires pour réaliser une cartographie du paysage gouvernemental du partage de données.

Constitution de l'équipe et de la coalition

Il est important de rassembler la bonne équipe pour le développement de l'initiative. Le développement d'une initiative de partage de données nécessite une collaboration inter-équipes dès le début.

L'équipe de base doit **inclure** des membres ayant de l'expérience et des responsabilités dans les domaines suivants :

- **Politique des données et de l'IA**, y compris la capacité de mener des analyses d'impact pour le partage de données et l'utilisation de systèmes d'IA dans le secteur public.
- **Expertise technique**, avec la capacité de générer des cas d'usage pour le partage de données, de discuter de la conception de l'architecture de données et de diriger sa mise en œuvre.
- **Gestion des parties prenantes** gouvernementales et **gestion de projet**.
- **L'expertise en données sectorielles** au sein du secteur dans lequel l'équipe espère partager des données sera également essentielle. Cependant, le secteur peut être déterminé à un stade ultérieur du projet et cette considération est couverte par la [question 3](#) de la liste de contrôle.

L'équipe de base doit **interagir** avec des représentants des groupes suivants :

- **Équipes juridiques** ayant de l'expérience dans l'application de cadres juridiques dans un contexte de partage de données et dans la rédaction d'accords de partage de données.
- **Instances de régulation** responsables de l'application des cadres de gouvernance des données du pays, ayant la capacité de s'engager fréquemment dans le projet et l'autorité d'assurer une supervision indépendante.
- **Équipes d'approvisionnement** ayant de l'expérience dans l'approvisionnement de logiciels et d'IA.
- **Équipe de communication** qui peut soutenir les activités d'engagement et de participation du public.

L'équipe de base doit avoir **accès** aux décideurs finaux au sein du département qui peuvent autoriser le mandat et le budget dont l'équipe a besoin pour avancer.

La participation de tous ces domaines et compétences dès le début facilite la communication aux points de décision et limite les conflits aux étapes ultérieures.

Assurer les capacités juridiques et réglementaires

Contrairement à de nombreux autres défis auxquels les équipes peuvent réagir plus rapidement avec un budget suffisant et le bon mandat, la présence **d'un cadre législatif et d'une capacité réglementaire sont des prérequis** pour aller de l'avant avec une initiative que l'équipe ne pourra pas résoudre seule.

Il doit exister un cadre législatif pour régir le partage de données. Cela inclut la législation sur la protection des données et la confidentialité, ainsi que d'autres législations qui établissent les limites et les exigences concernant le moment où les données peuvent être partagées, par qui, et à quelles fins. **Ces éléments sont détaillés plus en profondeur dans la cartographie du paysage du partage de données à la page suivante.**

En plus d'avoir une législation en place, une autre condition préalable est l'existence d'une agence de réglementation, responsable de la supervision de la mise en œuvre des cadres de gouvernance des données et, le cas échéant, de gouvernance de l'IA dans le pays. Au-delà de la simple création d'une agence de réglementation, celle-ci doit avoir (1) l'autorité pour faire respecter le cadre réglementaire afin qu'il ne puisse être ignoré par le gouvernement, (2) le savoir-faire pour appliquer la législation dans le contexte des initiatives de partage de données, et pour l'utilisation de données pour entraîner ou déployer des systèmes d'IA, et (3) la capacité de s'engager avec l'équipe et de donner des conseils pendant le développement et la mise en œuvre de l'initiative.

Par exemple, dans le cadre de la loi sur la protection de la vie privée de 2012 du pays (Republic Act No 10173, 2012), le Commissaire à la vie privée des Philippines est chargé, entre autres choses, de (1) veiller à la conformité des contrôleurs de données, y compris les agences gouvernementales; (2) publier des directives sur toutes les lois relatives à la protection des données; et (3) fournir une assistance à la demande des agences gouvernementales. Dans d'autres pays, les régulateurs développent également des directives et une expertise spécifiques à l'IA (OCDE, 2024). En France, son autorité de protection des données, la CNIL, a publié des recommandations pour les développeurs utilisant des ensembles de données d'entraînement pour l'IA qui incluent des données personnelles (CNIL, 2024).

Cartographie du paysage de partage de données

Une cartographie du paysage de partage de données décrit les principaux facilitateurs des initiatives de partage de données au sein du gouvernement (GPAI, 2023). Ce sont :

- l'infrastructure de données disponible pour soutenir une initiative de partage de données;
- les cadres politiques et législatifs qui facilitent et définissent les limites du partage de données; et
- les membres de l'équipe, les compétences et les relations pour faciliter la collaboration.



Recommandation des pilotes de la feuille de route

Lors des mises en œuvre pilotes de la feuille de route, les exercices de cartographie ont été menés sous forme d'ateliers avec l'équipe pilote de base et toutes les parties prenantes ayant une expérience ou des connaissances pertinentes afin de créer une image complète de l'état du partage de données au sein du gouvernement pilote. Les avantages de la cartographie du paysage de partage de données étaient de trois ordres.

1. Cela a favorisé l'alignement entre les membres de l'équipe et les autres parties prenantes concernant leurs **motivations pour entreprendre l'initiative et ses objectifs**.
2. Cela a permis aux équipes **d'identifier les facilitateurs du partage de données** au sein de leur gouvernement — par exemple, les infrastructures existantes et les relations avec les parties prenantes sur lesquelles l'initiative peut s'appuyer — ainsi que **les défis potentiels liés au partage de données** ; par exemple, le manque d'expérience au sein de l'équipe pour appliquer les cadres législatifs pertinents.
3. Elle a aidé à **la formation de coalitions** entre les équipes gouvernementales. Rassembler des parties prenantes de différentes équipes impliquées dans le partage de données peut aider à sensibiliser à l'initiative et à établir les relations nécessaires pour avancer.

Infrastructure de données

Plusieurs composantes sont pertinentes pour la cartographie de l'infrastructure de données (GPAI, 2023) :

- l'infrastructure utilisée pour le stockage, le traitement, l'intégration, la gestion, l'accès, la sécurisation et l'analyse des données ;
- les serveurs, qu'ils soient locaux ou basés sur le *cloud* ;
- les mécanismes d'accès aux données tels que les interfaces de programmation d'applications (API), les plateformes de données ouvertes du gouvernement ; et
- les politiques visant à établir des normes de données, l'interopérabilité et la qualité.



Paysage législatif

La législation actuelle et à venir qui est pertinente pour le partage de données dans la juridiction est importante. Bien que la législation pertinente diffère d'un pays à l'autre, six domaines législatifs clés sont importants pour garantir un accès équitable aux données par des développeurs d'IA tiers (GPAI, 2023) :

- **Protection et confidentialité des données** : protège les droits des personnes concernées³ et accorde des pouvoirs exécutoires aux institutions de supervision indépendantes pour réguler les systèmes d'IA.
- **Propriété intellectuelle (PI)** : les jeux de données peuvent être protégés par les lois sur la PI. Le partage de données peut donc avoir des implications sur le droit d'auteur ou les lois sur les secrets des affaires et peut nécessiter une licence de PI préalable. Il y a également des considérations sur les droits de PI sur les innovations issues de l'utilisation des données partagées.
- **Droit de la concurrence** : l'accès aux données peut créer un avantage économique pour les entreprises impliquées dans une initiative de partage de données.
- **Flux de données transfrontaliers** : la législation sur le maintien de la souveraineté des données peut influencer le partage de données avec des partenaires internationaux. Une telle législation peut imposer certaines exigences d'adéquation avant un transfert international. Ou, elle peut introduire des restrictions transfrontalières pour certains types de données dans certains secteurs.
- **Accès à l'information** : crée des mécanismes pour renforcer la transparence en donnant aux individus l'accès aux informations que le gouvernement détient sur eux.

Législation spécifique à l'intelligence artificielle : impose des exigences spécifiques au développement et au déploiement des systèmes d'IA, par exemple en fonction de leur niveau de risque (Commission Européenne, 2024).

³ La législation sur la protection des données et la confidentialité varie selon les juridictions, tout comme les protections qu'elle offre aux personnes concernées. Une explication des droits relatifs aux données est incluse dans les [Ressources supplémentaires](#) pour guider les équipes sur les points à prendre en compte dans la législation et les autres politiques.



Liste de contrôle pour le partage de données gouvernementales pour l'IA

Le guide de la liste de contrôle pour le partage de données gouvernementales pour l'IA est présenté aux pages suivantes. L'objectif de cette liste de contrôle est d'identifier les considérations que les équipes gouvernementales sont invitées à prendre en compte lorsqu'elles prennent une décision pour les trois étapes de décision du cadre :

- Cas d'usage (U)
- Mécanisme (M), et
- Structure de gouvernance (S).

Pour chaque étape de décision, les facteurs décisionnels présentés précédemment aident à cadrer la question posée.

Avant de prendre une décision à chaque étape, il est recommandé **que les équipes gouvernementales puissent répondre avec confiance "oui" à toutes les questions de la liste de contrôle.**

Les trois **sections** suivant la liste de contrôle (qui correspondent aux étapes de décision "Cas d'usage", "Mécanisme" et "Structure de gouvernance") détaillent chaque question et fournissent aux équipes un soutien pour :

1. évaluer si elles peuvent valider cette question dans la liste de contrôle, et
2. être en mesure de pouvoir valider cette question si ce n'est pas déjà le cas.

Le guide pour chaque question comprend :

- **Pourquoi c'est important** : la raison pour laquelle cette question est cruciale pour garantir que l'initiative de partage de données pour l'IA est responsable et efficace.
- **A quoi ça ressemble** : une description du moment où une équipe serait en mesure de cocher cette question dans la liste de contrôle. Pour la plupart des questions, cela diffère selon l'initiative, mais des exemples illustratifs sont fournis lorsque cela est possible.
- **Comment y arriver** : certaines questions incluent des activités suggérées pour aider l'équipe à satisfaire aux exigences de la question de la liste de contrôle. Des ressources d'appoint sont également fournies dans certains cas.
- **Recommandations découlant des pilotes de la feuille de route** : les enseignements tirés de la mise en œuvre de la feuille de route lors des projets pilotes qui ont été menés à l'aide de la version préliminaire sont parfois inclus.

Veuillez noter : Vous n'avez pas besoin de lire tout le contenu des encadrés bleus si vous estimez que vous comprenez la question et que vous pouvez y répondre "oui".

Le guide fourni pour chaque question est accessible par le biais des hyperliens dans la colonne de droite.

[Une version de la liste de contrôle au format tableur pour suivre les progrès peut également être téléchargée.](#)



Étape de décision : Cas d'usage (U)

Facteurs décisionnels	Question	Référence
Stratégie et durabilité	L'initiative cherche-t-elle à résoudre des problèmes clairs qui ont une valeur publique?	U1
Communauté et Parties prenantes	Existe-t-il une demande pour les données parmi les utilisateurs de données visés ?	U2
	L'initiative bénéficie-t-elle du soutien de la ou des équipes pertinentes responsables du secteur ou du domaine dans lequel les données sont partagées?	U3
	Le public soutient-il le partage de données dans ce secteur?	U4
Aspects juridiques et gouvernance	Existe-t-il un cadre juridique qui énonce clairement les conditions de partage des données applicables à cette initiative et respectées par celle-ci?	U5
	Les données à partager restent-elles dans les limites de ce qui est nécessaire pour le cas d'usage?	U6
Technologie	Les données partagées dans l'initiative peuvent-elles être préparées pour être utilisées par des développeurs d'IA?	U7
Éthique	Les applications d'IA prévues sont-elles conformes aux directives gouvernementales ou à d'autres meilleures pratiques en matière d'utilisation et de développement responsables de l'IA?	U8
	Les risques pour les personnes concernées et les autres groupes affectés peuvent-ils être atténués?	U9

Étape de décision : Mécanisme (M)

Facteurs décisionnels	Question	Référence
Stratégie et durabilité	Le modèle financier de l'initiative favorise-t-il une répartition équitable de ses bénéfices?	M1
Communauté et Parties prenantes	Les contrôleurs de données sont-ils d'accord sur la gestion de l'accès et de l'utilisation des données au sein de l'initiative?	M2
	Les responsables du traitement des données sont-ils en mesure et disposés à fournir des données pour l'initiative?	M3
Aspects juridiques et gouvernance	Existe-t-il des accords de partage de données entre toutes les parties nécessaires?	M4
Technologie	Existe-t-il une infrastructure de données qui supporte ce mécanisme de partage de données?	M5
	Est-il possible de partager les données de manière sûre et sécurisée?	M6
	Des pratiques de gestion et de gouvernance des données ont-elles été établies pour garantir la qualité continue des données tout au long de l'initiative?	M7
Éthique	Le niveau d'ouverture est-il proportionnel au niveau de sensibilité des données?	M8



Étape de Décision : Structure de gouvernance (G)

Facteurs décisionnels	Question	Référence
Stratégie et durabilité	La responsabilité de la définition, du suivi et de l'orientation de la stratégie globale de l'initiative est-elle attribuée?	G1
	Un budget est-il garanti pour cette initiative de partage de données, ou y a-t-il des options pour son financement?	G2
Communauté et Parties prenantes	Existe-t-il des mécanismes de communication des représentants de tous les acteurs affectés par l'initiative vers les personnes responsables de sa supervision?	G3
	La transparence est-elle maintenue tout au long du développement et de la mise en œuvre de l'initiative?	G4
	Existe-t-il un plan pour permettre à un large éventail d'acteurs d'accéder aux bénéfices de l'initiative?	G5
Aspects juridiques et gouvernance	L'initiative recueille-t-elle le consentement des personnes concernées par les données et leur permet-elle de le retirer à tout moment?	G6
	Existe-t-il un mécanisme par lequel les personnes concernées par les données peuvent demander réparation pour des violations de leurs droits?	G7
	Existe-t-il un canal pour que les agences de réglementation pertinentes puissent suivre l'initiative et donner des conseils sur son développement?	G8
	Existe-t-il des processus en place pour surveiller la conformité aux accords de partage de données?	G9
Technologie	Existe-t-il des processus en place pour surveiller l'infrastructure et maintenir un partage de données sûr et sécurisé pendant toute la durée de l'initiative?	G10
	Le gouvernement est-il équipé pour mener un processus d'approvisionnement équitable qui évalue les systèmes d'IA afin de s'assurer qu'ils répondent aux exigences de l'initiative?	G11
Éthique	Existe-t-il des processus pour surveiller et gérer les risques pour les personnes concernées par les données et les autres groupes affectés?	G12
	L'initiative exige-t-elle une surveillance humaine pour toute prise de décision algorithmique?	G13



Étape de décision : Cas d'usage

Dans la feuille de route, les cas d'usage de partage de données pour les initiatives d'IA se caractérisent par :

- **Un problème ou un ensemble de problèmes cibles** : le(s) problème(s) que l'équipe souhaite aborder à l'aide d'une solution qui implique le partage de données.
- **Les données à partager** : les données qui doivent être partagées pour aborder ce problème.
- **Les applications d'IA** : la manière dont les applications d'IA peuvent aider à résoudre le problème cible en utilisant les données.
- **Les utilisateurs de données** : les types de parties prenantes (ou des parties prenantes spécifiques si elles sont connues) qui utiliseront les données partagées pour développer des applications d'IA.

Pourquoi cette étape de décision est importante :

Pour s'assurer que l'initiative est motivée par la résolution de problèmes concrets d'intérêt public, il est nécessaire d'identifier et de sélectionner des cas d'usage clairs. L'équipe a besoin d'un cadre pour identifier les cas d'usage en fonction de la demande de données, qui appliquent des cadres juridiques et de gouvernance clairs définissant les limites du partage de données et qui œuvrent dans le sens des priorités gouvernementales.

Comment utiliser cette section de la liste de contrôle :

Ces questions aident les équipes à identifier, évaluer et prioriser les cas d'usage pour l'initiative. La priorisation peut être nécessaire en raison de contraintes de ressources qui obligent les équipes à choisir entre plusieurs cas d'usage possibles.

Cette section de la liste de contrôle peut également servir de cadre d'évaluation quantitative pour comparer et prioriser les cas d'usage potentiels. Un modèle est inclus dans la section [Ressources supplémentaires](#), ainsi qu'une matrice de priorisation.

Si l'équipe utilise la liste de contrôle pour identifier des cas d'usage et n'a pas encore de secteur ou de domaine problématique spécifique en tête, il est recommandé de d'abord délimiter l'espace du problème à un secteur, par exemple en se basant sur les projets existants ou les priorités gouvernementales.

Cas d'usage – Stratégie et durabilité

U1. L'initiative cherche-t-elle à résoudre des problèmes clairs qui ont une valeur publique?

Pourquoi c'est important

Le succès de l'initiative dépend du partage de données qui sont précieuses pour les acteurs qui souhaitent résoudre un problème auquel ils sont confrontés, ou auquel d'autres personnes sont confrontées. Par conséquent, les cas d'usage de l'initiative doivent identifier et cibler des problèmes particuliers.

Une façon de déterminer si un problème a une valeur publique est de considérer les priorités gouvernementales dans des secteurs, tels que la santé, ou au sein du gouvernement, comme la numérisation des services publics. En se concentrant sur un problème qui est déjà une priorité gouvernementale, il est plus facile d'assurer et l'adhésion des dirigeants et d'obtenir un financement.

À quoi cela pourrait ressembler

Le ou les problèmes ciblés peuvent être ceux auxquels font face le gouvernement, les acteurs privés, les chercheurs ou le grand public. Par exemple, la plateforme colombienne Aclímate est une réponse aux problèmes auxquels les agriculteurs sont confrontés en raison du changement climatique. Plus précisément, leurs rendements souffrent de schémas météorologiques changeants et de plus en plus imprévisibles. Les informations issues de l'apprentissage automatique (machine learning) et axées sur les données les aident désormais à gérer ces changements (Young et Verhulst, 2017).

Comment y arriver

L'identification des cas d'usage impliquera de travailler avec des acteurs internes et externes au gouvernement pour comprendre où un manque de données entrave les progrès dans leur travail. En travaillant avec des acteurs pour trouver des cas d'usage potentiels, l'équipe peut adopter une approche ascendante (*bottom-up*), descendante (*top-down*) ou mixte pour exploiter les connaissances du domaine des acteurs sectoriels :

- **Une approche ascendante** invite des propositions de multiples acteurs – qu'ils soient gouvernementaux ou non-gouvernementaux – par exemple, par le biais d'appels ouverts pour présenter des idées de partage de données. Cette approche peut être moins intensive en ressources et aider à créer des partenariats avec des acteurs clés. Cependant, elle nécessite une sensibilisation efficace, une volonté d'engagement de la part des acteurs et peut limiter le nombre de cas d'usage potentiels identifiés.
- **Une approche descendante** implique de mener des ateliers, des enquêtes, des entretiens et des recherches documentaires sur les défis auxquels sont confrontés les acteurs travaillant dans le secteur, tant à l'intérieur qu'à l'extérieur du gouvernement, pour identifier où il y a une demande, une offre de données et un précédent. Cette approche peut aider à identifier un éventail de cas d'usage potentiels, ce qui permet de hiérarchiser le plus approprié. Cependant, elle peut être plus intensive en ressources s'il n'y a pas de demandes existantes pour un partage de données accru de la part des acteurs du secteur.

Les processus utilisés pour engager les acteurs pendant le développement de l'initiative doivent être ouverts et équitables afin que tout futur fournisseur potentiel du gouvernement ne bénéficie pas d'un avantage indu. Cela pourrait inclure la tenue de consultations publiques ou d'appels ouverts à l'engagement.

Recommandation des pilotes de la feuille de route.

Pour identifier les acteurs à impliquer et à explorer les problèmes ciblés possibles, il est utile de réaliser une cartographie de l'écosystème de données existant. Une carte de l'écosystème de données visualise quels actifs de données existent dans un écosystème, et comment ils sont consultés, utilisés et partagés entre les acteurs (ODI, 2022).

Une fois la carte de l'écosystème de données terminée, elle peut être utilisée pour explorer ou valider les opportunités de partage de données et pour visualiser les obstacles actuels auxquels sont confrontés les acteurs, ainsi que les points où l'initiative pourrait soutenir de nouveaux flux de données et le développement et le déploiement de systèmes d'IA au sein de l'écosystème.



L'ODI a créé une [méthodologie](#) (ODI, 2022) et des [outils](#) (ODI, 2019) pour réaliser une cartographie de l'écosystème de données. Cette cartographie est de préférence réalisée avec la participation des acteurs concernés.

Cas d'usage – Communauté et Parties prenantes

U2.Existe-t-il une demande pour les données parmi les utilisateurs de données visés?

Pourquoi c'est important

S'il n'y a pas d'acteurs qui souhaitent utiliser les données partagées pour travailler sur le problème cible, les données ne seraient pas utilisées pour créer de la valeur publique.

À quoi cela pourrait ressembler

Il existe une demande claire et démontrable pour les données liées à ce problème de la part d'acteurs de l'industrie, du milieu universitaire, de la société civile et du gouvernement. Précisément, il existe des preuves que les restrictions d'accès aux données bloquent actuellement des applications d'IA potentielles et des avancées à l'intérieur ou à l'extérieur du gouvernement.

Exemples de demande de données traitée par des initiatives de partage de données

Initiative (Pays)	La demande de données de la part des acteurs
Aclímate (Colombia)	Les chercheurs et les prévisionnistes ont besoin de données précises, complètes et en temps réel pour développer des analyses sur les schémas climatiques et environnementaux. Les agriculteurs utilisent ces informations pour guider leurs décisions en matière de pratiques agricoles et de stratégies de plantation optimales afin de maximiser les rendements. La collaboration entre l'IDEAM, le CIAT et la Federroz facilite le partage et l'intégration sécurisés des données du gouvernement et des agriculteurs sur la plateforme Aclímate. Ces données sont ensuite complétées par des outils et des informations d'apprentissage automatique qui permettent aux agriculteurs d'accéder à des informations exploitables (Young et Verhulst, 2017).
The Health Passbook (Taiwan)	Les applications tierces qui fournissaient des services de santé personnalisés et celles qui menaient des activités de recherche et développement pour les technologies d'IA de santé n'avaient pas d'accès direct aux données de la NHIA ou des patients. Compte tenu de la nature sensible des données, ils ne pouvaient pas accéder aux serveurs, ce qui rendait la fourniture de leurs services plus difficile. Un mécanisme a été mis en place pour permettre aux patients de consentir à ce que leurs données soient utilisées par des applications tierces validées par la NHIA. Les patients pouvaient autoriser directement les développeurs d'applications d'IA tierces et les chercheurs, qui avaient été examinés par la NHIA, à accéder à leurs données pour fournir des services supplémentaires (y compris la recherche pour l'IA de santé) (GPAI, 2023).
Copernicus (EU)	En l'absence de données satellitaires et géospatiales facilement accessibles et précises, les analystes gouvernementaux, les chercheurs et les scientifiques ne peuvent pas surveiller adéquatement l'environnement, gérer les risques de catastrophes ou mener des exercices de santé publique. Une pénurie de données complètes et à jour entrave



	l'utilisation d'outils d'IA puissants pour l'analyse des données climatiques. Le Copernicus Open Access Hub fournit des données opportunes, bien structurées et traitées sur le climat, la géographie et l'environnement, permettant aux utilisateurs de mener des analyses approfondies et d'employer des outils d'IA (Copernicus, n.d.).
All of Us data (USA)	Les chercheurs et les responsables de la santé publique ont besoin de données de santé complètes et représentatives de l'ensemble de la population pour éclairer des décisions politiques de santé robustes. L'accès aux données au niveau du patient est essentiel pour la recherche et le développement de la médecine personnalisée. En particulier, les grands jeux de données sont vitaux pour tirer parti de l'apprentissage automatique et des outils d'IA afin d'obtenir de nouvelles informations et d'éclairer les modèles prédictifs sur la santé de la population, la gestion des maladies et les résultats des traitements. Auparavant, l'accès à ces données était difficile en raison d'incohérences dans les méthodes de traitement et de nettoyage, qui ne respectaient pas adéquatement les droits à la vie privée, ainsi que d'une accessibilité limitée. La plateforme All of Us offre un moyen sécurisé d'accéder à ces données (All of Us, n.d.).
Land Transport Authority (LTA) DataMall (Singapore)	Des données complètes en temps réel et historiques sont nécessaires pour faciliter une bonne planification du transport urbain et permettre la recherche en IA pour améliorer les embouteillages et les flux de trafic. Les développeurs d'applications ont également besoin d'accéder à des informations de trafic en temps réel pour fournir des mises à jour et des informations en temps réel aux citoyens utilisant ces services. Ces données de transport, cependant, proviennent de diverses sources et seraient difficiles d'accès. La plateforme DataMall de Singapour agrège et centralise ces données pour fournir un point d'accès unique à un large éventail de jeux de données. La plateforme fournit des API claires qui permettent un accès varié aux jeux de données sur demande. Elle permet aux chercheurs universitaires, aux entreprises du secteur privé, aux autres départements gouvernementaux et aux citoyens ordinaires d'accéder à toutes les données de transport pertinentes en un seul endroit (Land Transport Authority, n.d.).

U3. L'initiative bénéficie-t-elle du soutien de la ou des équipes pertinentes responsables du secteur ou du domaine dans lequel les données sont partagées? *[Communauté et Parties prenantes]*

Pourquoi c'est important

Le soutien de l'équipe responsable du secteur ou du domaine, si ce n'est pas l'équipe qui mène l'initiative, est essentiel au succès de celle-ci. L'équipe sectorielle est susceptible d'être l'actuelle curatrice et contrôleur des données pertinentes pour le secteur ou le problème. Elle est également susceptible d'avoir des relations avec les parties prenantes pertinentes en dehors du gouvernement.

À quoi cela pourrait ressembler

Si l'initiative n'est pas menée par une équipe qui est responsable du secteur ou du domaine, l'initiative bénéficie de l'engagement de la ou des équipes sectorielles. Par exemple, si le projet est géré par une équipe du gouvernement numérique qui travaille dans l'ensemble du gouvernement et souhaite soutenir le partage de données dans le secteur agricole, alors l'équipe devra collaborer étroitement avec l'agence ou l'équipe responsable de l'agriculture.



U4. Le public soutient-il le partage de données dans ce secteur? [Communauté et Parties prenantes]

Pourquoi c'est important

Le soutien du public est fondamental pour établir la confiance dans l'initiative. Il est d'autant plus important de déterminer si le public soutient le partage de données lorsque des données personnelles, ou d'autres données très sensibles, comme celles relatives à la santé ou à la sécurité nationale, sont partagées.

À quoi cela pourrait ressembler

Des indicateurs issus de recherches, telles que des enquêtes, des études sur les utilisateurs et des consultations, ou d'initiatives gouvernementales précédentes, montrent que le public fait confiance au gouvernement pour permettre le partage de données dans ce secteur.

Cas d'usage – Aspects juridiques et gouvernance

U5. Existe-t-il un cadre juridique qui énonce clairement les conditions de partage des données applicables à cette initiative et respectées par celle-ci?

Pourquoi c'est important

Les cas d'usage que l'initiative soutient doivent être conformes à la législation et aux orientations réglementaires sur le partage de données.

À quoi cela pourrait ressembler

Les cadres juridiques et les orientations réglementaires différeront d'une juridiction à l'autre. Dans toutes les juridictions, les équipes doivent identifier les cadres, et les orientations qui les accompagnent, qui se rapportent aux cas d'usage. En plus de la législation sur la protection et la confidentialité des données, il peut y avoir une législation plus spécifique qui doit être respectée selon le cas d'usage; par exemple, cela peut être le cas si des données de santé sont partagées.

En utilisant ces cadres, et avec les conseils de membres de l'équipe ayant une expertise juridique, les équipes devront établir les différentes autorisations et exigences pour le partage de données dans le cadre de leur initiative et les appliquer aux cas d'usage à l'étude. Celles-ci peuvent inclure⁴ :

- Si l'organisme gouvernemental a le pouvoir légal de partager des données
- Si l'initiative a besoin d'autorisations pour partager des données
- S'il existe des bases légales spécifiques pour le partage des données qui doivent être établies
- Les fins pour lesquelles l'organisme peut partager des données
- Toute documentation spécifique qui doit être conclue
- Toute exigence relative aux mesures de confidentialité et de sécurité que l'équipe doit prendre

⁴ Il s'agit d'une liste illustrative des types d'exigences que les initiatives doivent identifier et appliquer aux cas d'usage qu'elles souhaitent partager dans le cadre de leurs projets. Ces exigences varient selon les juridictions. Des exemples internationaux sont fournis pour illustrer la diversité des exigences entre les différents pays.



Exemples internationaux de différentes exigences

Législation (Pays)	Exemple de conditions relatives au partage de données
Loi Canadienne sur l'Autorité Légale (Canada's Legal Authority Requirement)	Pouvoir juridique de partager des données : une législation spécifique à un programme ou à un ministère précise si une initiative a l'autorité légale de collecter et d'utiliser des données personnelles (Gouvernement du Canada, 2023).
Loi Britannique sur l'Économie Numérique (United Kingdom's Digital Economy Act)	Permissions pour le partage de données : un organisme de traitement, qui partage des données à des fins de recherche, doit être accrédité par l'Autorité statistique du Royaume-Uni avant de pouvoir le faire (Loi sur l'économie numérique, 2017).
Règlement général sur la protection des données (RGPD) de l'Union européenne (European Union's General Data Protection Regulation (GDPR))	Bases légales : le traitement des données personnelles n'est licite que pour un ensemble de bases légales spécifiées, qui incluent le consentement des personnes concernées, les obligations contractuelles et l'intérêt légitime d'une organisation (Règlement 2016/679).
Loi Turque sur les Statistiques (Turkiye's Statistics Law)	Objectifs du partage de données : les données au niveau individuel peuvent être partagées pour être utilisées dans des études scientifiques sans aucune référence à des unités statistiques distinctes (Loi sur les statistiques de la Turquie, 2005).
Loi Indonésienne sur la Protection des Données Personnelles (Indonesia's Personal Data Protection Law)	Documentation spécifique : une évaluation d'impact sur la protection des données doit être réalisée pour le traitement des données personnelles classées comme à haut risque (Dharyanto, Murata et Ustriyana, 2023).
Mesures Chinoises relatives à l'Information sur la Santé de la Population (China's Population Health Information Measures)	Exigences de confidentialité et de sécurité : exige que les données sur la santé de la population ne soient pas stockées ou hébergées sur des serveurs en dehors de la Chine et établit des exigences de stockage de manière graduée en fonction de la sensibilité des données (Postigo, 2023).

Comment y arriver

L'équipe peut s'engager avec des représentants des équipes juridiques et réglementaires pour établir comment le cadre juridique s'applique à leur initiative, y compris les conditions et les limites du partage de données qui sont pertinentes.

U6. Les données à partager restent-elles dans les limites de ce qui est nécessaire pour le cas d'usage? [*Aspects juridiques et Gouvernance*]

Pourquoi c'est important

Le partage uniquement des données nécessaires à des fins spécifiées et légitimes contribue à garantir que les données ne sont utilisées que pour résoudre le problème spécifié et que le cas



d'usage respecte le principe de proportionnalité (Black and Stevens, 2013). Ceci est important pour aider à prévenir le "détournement d'usage" (ou *function creep*), où les données sont traitées à des fins qui vont au-delà de ce qui était initialement prévu (Koops, 2021). La spécification des finalités dépend du type et de la sensibilité des données partagées. Il peut être possible de publier des données sous une licence de données ouvertes, auquel cas les finalités seront illimitées. Ce sera différent pour les données personnelles ou les données sensibles non personnelles, où les finalités doivent être définies et les données partagées restreintes conformément à ces finalités.

À quoi cela pourrait ressembler

Dans les cas où les données ne sont pas partagées sous une licence de données ouvertes, les données partagées ne dépassent pas ce qui est nécessaire pour atteindre l'objectif légitime du cas d'usage.

Cas d'usage – Technologie

U7. Les données partagées dans l'initiative peuvent-elles être préparées pour être utilisées par des développeurs d'IA?

Pourquoi c'est important

Les données doivent être facilement accessibles pour les utilisateurs. Elles doivent donc être détenues/stockées/formatées de manière à répondre aux besoins des utilisateurs potentiels, tels que les développeurs d'IA, afin qu'ils puissent travailler sur le problème défini par le cas d'usage.

À quoi cela pourrait ressembler

De meilleures pratiques de gestion des données et l'application de normes de données reconnues par les contrôleurs de données sont des indicateurs que les données sont susceptibles d'être utiles. Un premier indicateur supplémentaire est le fait que les personnes qui utilisent actuellement les données les considèrent comme fiables. Cependant, les exigences spécifiques des données pour les futurs utilisateurs doivent être déterminées et évaluées. Cela peut concerner le format et la structure des données, en veillant par exemple à ce qu'elles soient lisibles par une machine.

Comment y arriver

Comprendre et satisfaire les besoins des utilisateurs en matière de données impliquera de consulter les utilisateurs potentiels et d'évaluer les données qui seront partagées par rapport à leurs exigences.

Évaluation des données : Comment rendre les données exploitables par l'IA ?

L'objectif d'une évaluation des données

Quatre facteurs principaux doivent être pris en compte lors de l'évaluation de la pertinence et de l'utilité d'un ensemble de données pour un cas d'usage :

- **Qualité et exhaustivité** : évalue la précision et l'exhaustivité de l'ensemble de données lui-même.



- **Pertinence pour le cas d'usage** : évalue dans quelle mesure les données correspondent aux besoins du cas d'usage particulier.
- **Éthique et conformité légale** : évalue si l'ensemble de données a été créé de manière appropriée et s'il peut être mis en conformité avec les normes de confidentialité et de sécurité pertinentes.
- **Besoins de (pré)traitement** : évalue dans quelle mesure les données sont formatées pour le cas d'usage particulier et si un nettoyage supplémentaire est nécessaire.

Une analyse complète de ces dimensions, y compris des questions pour l'équipe afin qu'elle puisse mener sa propre évaluation, est fournie dans la section [Ressources complémentaires](#)

Il est important de noter que l'objectif de cette évaluation est de s'assurer que les données répondent aux besoins des utilisateurs. Elle est distincte d'une analyse d'impact sur les données, qui peut être également requise dans le cadre de l'établissement de la dimension éthique et de conformité légale.

L'adaptation de l'évaluation des données

Les caractéristiques de l'évaluation doivent être adaptées aux besoins spécifiques du cas d'usage. Différents types de données utilisées dans des contextes différents nécessitent des normes de précision différentes, impliquent des besoins de conformité réglementaire et éthique distincts, et requièrent des formats variés. Il est utile de considérer ici :

- **le type (et la sensibilité) des données utilisées** : s'agit-il de données personnelles, de données de trafic, de données géospatiales ?
- **l'utilisation particulière dans le développement de l'IA** : sont-elles utilisées comme ensemble d'entraînement, données de test, données d'entrée ?
- **le type de techniques d'IA utilisées** : sont-elles utilisées pour l'apprentissage supervisé, non supervisé, l'IA générative ?
- **le but du cas d'usage** : sont-elles utilisées pour éclairer une politique, prendre des décisions, générer des informations ?

Réalisation d'une évaluation des données

Il est nécessaire de s'assurer, avec les participants potentiels de l'initiative, que les ensembles de données sont appropriés pour le cas d'usage. L'évaluation des données [fournies](#) sert de cadre pour organiser un atelier avec les parties prenantes concernées par le cas d'usage. Les acteurs responsables de la collecte, du traitement et de l'utilisation des données doivent être présents, ainsi que les futurs utilisateurs potentiels des données fournies par le gouvernement. Des données fictives doivent être créées si nécessaire pour partager et évaluer les propriétés des ensembles de données pertinents avec des parties externes.



Cas d'Usage – Éthique

U8. Les applications d'IA prévues sont-elles conformes aux directives gouvernementales ou à d'autres meilleures pratiques en matière d'utilisation et de développement responsables de l'IA ?

Pourquoi c'est important

L'initiative doit aider à promouvoir le développement et l'utilisation de l'IA qui apportent des avantages publics, respectent les valeurs d'égalité et d'équité, et ne causent pas de préjudice aux individus.

À quoi cela pourrait ressembler

Si l'équipe envisage un cas d'usage dans lequel des données sont partagées pour le déploiement d'applications d'IA au sein du secteur public, alors l'équipe utilise des directives sectorielles ou intergouvernementales sur l'utilisation de l'IA dans le secteur public, ou des principes éthiques pour l'utilisation de l'IA dans le secteur public. S'il n'y a pas de ressources gouvernementales à ce sujet, la priorité est que l'équipe soit en mesure d'examiner les risques et les implications liés à son initiative.

Si l'équipe envisage des cas d'usage pour le développement et le déploiement de systèmes d'IA en dehors du gouvernement, par exemple à des fins de recherche ou commerciales, alors l'équipe peut réfléchir à la manière de guider la direction de ces applications en dehors du gouvernement. Par exemple, l'NHIA de Taiwan vérifie les développeurs d'applications tiers qui peuvent accéder aux données de santé des patients et du gouvernement via son Health Passbook. De même, les chercheurs qui demandent l'accès aux données du Data Service du gouvernement britannique doivent soumettre un formulaire de demande, détaillant comment leur projet vise à apporter des bénéfices publics (UK Data Service, n.d).

Outils disponibles

S'il n'existe pas de politiques gouvernementales établies pour l'utilisation de l'IA dans le secteur public, des outils existants peuvent être consultés, notamment la [Boîte à outils pour l'IA dans le secteur public](#) de l'UNESCO et de l'OCDE (UNESCO et OCDE, 2024), ou le [Trustworthy AI Self Assessment](#) d'Oxford Insights (Oxford Insights, 2023).

U9. Les risques pour les personnes concernées et les autres groupes affectés peuvent-ils être atténués? [Éthique]

Pourquoi c'est important

Chaque cas d'usage comportera un ensemble unique de risques survenant à différents moments du partage et de la réutilisation des données, et affectant différents groupes d'acteurs. Certains des risques posés par le cas d'usage peuvent ne pas avoir des conséquences directes. Par exemple, cela inclut des impacts sociétaux plus larges sur les groupes ayant accès aux données et ceux qui n'en ont pas, ou comment les inégalités numériques pourraient être amplifiées.

À quoi cela pourrait ressembler

Il existe un registre des risques ou une autre documentation qui enregistre et traite les risques qui surviennent à toutes les étapes de l'initiative. Les risques sont interprétés au sens large et



incluent les risques liés aux caractéristiques des données partagées, aux systèmes d'IA utilisés et aux implications sociales des applications d'IA.

Comment y arriver

Les risques peuvent être identifiés à l'aide de techniques telles que les analyses d'impact sur les données, les analyses d'impact sur l'IA et l'analyse des conséquences. Il peut déjà exister des pratiques établies au sein du gouvernement.

Outils Disponibles

S'il n'existe pas de politiques gouvernementales établies pour la réalisation d'évaluations des risques, il existe des outils pour soutenir ces techniques, comme le [Data Ethics Canvas](#) de l'ODI (ODI, 2021), l'outil [Consequence Scanning](#) de Doteveryone (Doteveryone, 2021) et [l'Évaluation de l'Impact éthique de l'IA](#) de l'UNESCO (UNESCO, 2023).

À ce stade, il est important d'examiner s'il existe des options pour répondre efficacement à un risque. En effet, tous les risques ne sont pas gérables et certains peuvent empêcher l'équipe de poursuivre le cas d'usage ; par exemple, si un cas d'usage risque de porter atteinte à des droits fondamentaux. Ces options d'atténuation peuvent être développées plus en détail dans la section sur la structure de gestion et de gouvernance.

Étape de décision : Mécanisme

Le mécanisme définit la manière dont les décisions concernant l'accès aux données et leur utilisation au sein de l'initiative seront prises et par qui. Ce faisant, il détermine l'infrastructure juridique et de données nécessaire pour mettre en œuvre ces décisions.

Pourquoi cette étape de décision est importante :

S'accorder sur le mécanisme de l'initiative est important afin d'établir la confiance entre les participants (Frontier Economics, 2021). Cela se fait en clarifiant :

- **Les rôles des participants** : il établit qui a le contrôle des données à différentes étapes de l'initiative et comment les décisions sont prises concernant ce contrôle.
- **Les accords juridiques** : il identifie les accords de partage de données nécessaires pour permettre la circulation des données.
- **Les exigences en matière d'infrastructure de données** : il définit les exigences de l'infrastructure de données nécessaire pour fournir l'accès aux données et garantir la confidentialité et la sécurité.

De plus, le mécanisme est important car il a des implications sur les cas d'usage possibles au sein de l'initiative et sur la complexité de la structure de gouvernance.

Comment utiliser cette section de la liste de contrôle:

La taxonomie des mécanismes de partage de données, fournie dans la section [Annexe 1 – Supporting Resources](#), donne des exemples de la manière dont le partage de données peut être organisé en fonction du nombre de participants impliqués et des flux de données souhaités entre eux.

Si l'équipe n'a pas encore décidé du type de mécanisme de partage de données qu'elle compte utiliser, la taxonomie peut être consultée pour identifier des exemples qui permettraient les flux de données dont elle a besoin au sein de l'initiative. Cette partie de la liste de contrôle peut être utilisée pour spécifier les rôles des participants, l'infrastructure de données et l'infrastructure juridique, afin que le mécanisme soit adapté à l'initiative et au contexte gouvernemental.

Si l'équipe a déjà décidé d'un mécanisme général, cette partie de la liste de contrôle peut être utilisée pour clarifier les rôles et examiner l'infrastructure juridique et de données requise ou existante qui soutient l'initiative.

Mécanisme – Stratégie et durabilité

M1. Le modèle financier de l'initiative favorise-t-il une répartition équitable de ses bénéfices?

Pourquoi c'est important

Le modèle financier de l'initiative définit qui la finance et détermine donc en partie comment sa valeur économique est répartie entre les participants. Il est important de considérer l'équité de cette répartition.

À quoi cela pourrait ressembler

L'initiative peut être financée, entièrement ou partiellement, par des fonds publics :

- **Financement public** : l'ensemble de l'initiative est financé par le gouvernement ou par des sources externes, comme une banque de développement.
- **Financement public-privé** : le financement public de l'initiative est complété par la facturation de l'utilisation des données, des services de gestion de données, ou des produits et services créés à partir de ces données (Ada Lovelace Institute, 2021a).

Comment aborder l'équité dans le modèle financier du mécanisme de partage de données :

(1) Le modèle financier reflète-t-il la valeur économique créée?

Cela signifie que le modèle met en balance le coût de la participation et la valeur économique pour les participants. Si l'équipe facture la participation, il est important de considérer comment la valeur économique est générée, pour qui et à quelle échelle, afin de s'assurer que les participants paient un coût qu'ils sont prêts à assumer.

Par exemple, les initiatives de données peuvent créer de la valeur économique pour les utilisateurs en leur permettant de créer des produits ou des services qu'ils vendront par la suite. L'échelle de la valeur économique dans ce cas dépend de l'utilisateur ; certains seront plus aptes que d'autres à transformer les données en gain financier. L'échelle dépend également de la qualité des données partagées. Si elles ne sont pas partagées dans un format facilement réutilisable, l'utilisateur peut avoir à investir dans leur préparation (GPAI, 2023). Des frais d'accès équitables tiendront compte de ces dépendances.

(2) Comment le modèle affecte-t-il la répartition des bénéfices?

Cela signifie que le modèle répartit les bénéfices afin de maximiser la valeur publique créée par l'initiative. Le modèle financier détermine en partie la répartition des bénéfices en affectant qui a accès aux données ainsi qu'aux produits et services créés à partir de ces données. Toute facturation pourrait entraîner l'exclusion de certains acteurs. Il est donc important de déterminer quels acteurs pourraient être exclus et quelles pourraient être les conséquences de cette exclusion.



L'équipe peut encourager un large accès aux bénéfices des produits et services créés à partir des données. Elle peut envisager d'imposer des conditions à son partenariat pour encourager les partenaires à utiliser les données dans l'intérêt public (GPAI, 2023).

L'élaboration d'un ensemble de principes que l'équipe souhaite refléter dans son modèle financier peut aider à guider sa décision. Cette approche a été adoptée par le Service national de santé du Royaume-Uni (NHS England, 2023).

Mécanisme – Communauté et Parties prenantes

M2. Les contrôleurs de données sont-ils d'accord sur la gestion de l'accès et de l'utilisation des données au sein de l'initiative?

Pourquoi c'est important

La gestion de l'accès et de l'utilisation des données consiste notamment à établir qui a le contrôle des données aux différentes étapes de l'initiative. Elle implique de définir comment les décisions sont prises et par qui (l'individu, l'organisation ou le groupe d'organisations responsables). L'établissement des rôles et des responsabilités clairs pour les participants tout au long de l'initiative est important pour générer la confiance entre eux, car les données sont partagées avec de nouvelles parties et à de nouvelles fins que les participants peuvent percevoir comme un risque.

À quoi cela pourrait ressembler

Tous les groupes d'acteurs se sont mis d'accord sur la manière dont les décisions sont prises et par qui. Les conséquences de ces décisions sur le contrôle des données sont également comprises par tous les groupes d'acteurs.

Cela peut être aussi simple qu'une autorité gouvernementale ayant le contrôle sur toutes les données nécessaires à un cas d'usage et transférant ce contrôle à une autre partie. Cependant, dans d'autres cas, il peut y avoir plusieurs contrôleurs de données, des contrôleurs intermédiaires ou plusieurs utilisateurs de données. Chacun de ces acteurs aura des exigences et des préoccupations différentes concernant la manière dont le contrôle des données évolue tout au long de l'initiative.

Définir les conditions de l'initiative de partage de données

Lorsque l'équipe met en place une initiative pour permettre un partage continu des données entre les participants, répondre à cette question implique également de définir les conditions de participation. Par exemple, si l'équipe met en place un "data commons" (un bien commun de données), elle devra établir les règles d'adhésion. Alternativement, si l'équipe utilise un modèle de gestion de données ("data stewardship"), elle devra établir les critères d'accès aux données. Les équipes peuvent se poser les questions suivantes :

Qui peut devenir un fournisseur de données?



- Qualité des données : respectent-elles les normes de qualité et de formatage des données ?
- Protection des données et confidentialité : sont-elles conformes aux exigences pertinentes ?

Qui peut devenir un utilisateur de données ?

- But de l'utilisation : est-ce, par exemple, pour la recherche ou des critères spécifiques d'intérêt public ?
- Conditions d'utilisation : comment les données sont-elles accessibles ; par exemple, est-ce uniquement via un lieu physique ?
- Obligations de surveillance de l'utilisation et des résultats.

Comment chaque demande sera-t-elle évaluée ?

- Le processus de prise de ces deux décisions concernant la participation à l'initiative.

Exemple : INSIGHT (Royaume-Uni)

INSIGHT est un fonds de données ("data trust") géré par le Service national de santé (NHS) du Royaume-Uni pour permettre le partage responsable de données de santé oculaire très sensibles afin de faciliter la recherche. INSIGHT utilise un processus d'approbation en trois étapes pour les demandes d'accès aux données (INSIGHT, n.d.) :

- *Sélection initiale* : s'assure que la recherche respecte le cadre des "cinq sécurités" (*five safes*).
- *Évaluation éthique* : évalue l'objectif, la valeur et les bénéfices de la recherche. Les critères incluent l'évaluation de l'atténuation des risques, la consultation publique et la garantie que le projet ne désavantage aucun groupe.
- *Approbation finale* : accordée par INSIGHT sur la base des recommandations du Comité consultatif du fonds de données (DataTAB). Le DataTAB se réunit tous les 2 mois pour discuter des demandes.

Comment y arriver

Il est essentiel que tous les participants soient d'accord sur la manière dont l'accès aux données est utilisé et géré pour qu'ils soient disposés à prendre part à l'initiative. Par conséquent, un accord doit être conclu de manière inclusive et impliquer l'engagement de tous les participants (Ada Lovelace Institute, 2021b).

Recommandation des projets pilotes de feuille de route

Il peut être utile de s'appuyer sur les responsabilités existantes des organisations et les méthodes de travail entre elles. En effet, la négociation de nouvelles responsabilités et l'établissement de nouvelles pratiques de travail nécessitent un engagement important et sont plus difficiles à négocier.

Par exemple, si les décisions concernant l'accès aux données doivent être prises collectivement, il peut être plus simple de les prendre au sein de forums existants. Alternativement, si la prise de décision est déléguée à une seule organisation, il est approprié



de choisir une organisation qui a déjà la confiance des participants pour exercer des fonctions similaires.

M3. Les responsables du traitement des données sont-ils en mesure et disposés à fournir des données pour l'initiative? *[Communauté et Parties Prenantes]*

Pourquoi c'est important

L'initiative imposera des exigences aux contrôleurs de données, qui peuvent inclure la préparation et la publication des données, la négociation des accords de partage de données, la gestion des demandes d'accès aux données. Il est important d'établir des exigences qui correspondent aux attentes des contrôleurs en termes de temps disponible, de capacité numérique, de ressources et de volonté de participer.

À quoi cela pourrait ressembler

Les rôles des contrôleurs de données actuels au sein de l'initiative sont clairs et convenus collectivement.

Recommandation des projets pilotes de feuille de route

L'initiative est susceptible de créer de nouvelles responsabilités pour les contrôleurs de données. En discutant avec les contrôleurs de données potentiels de leur capacité à assumer ces responsabilités, l'équipe peut adapter le rôle des contrôleurs de données à leurs besoins. Ceci est particulièrement important si l'initiative cherche à attirer des contrôleurs de données pour fournir des données.

Exemple : Service de données intégrées (IDS) (Royaume-Uni)

Le Service de données intégrées (IDS) du Royaume-Uni, géré par l'Office for National Statistics (ONS), facilite le partage de données sur la société et l'économie, provenant des départements gouvernementaux, avec des chercheurs accrédités.

L'IDS fournit à ses utilisateurs des jeux de données interreliés à l'échelle du gouvernement., mais il propose 3 modèles pour la manière dont les départements gouvernementaux peuvent fournir et lier les données afin de répondre à leurs différents besoins (Office for National Statistics, 2023) :

- *Modèle de don* : l'équipe de liaison de données de l'ONS ingère et lie les données pour le compte du fournisseur de données.
- *Modèle de dépôt* : il existe une zone de transit à l'intérieur de la plateforme IDS qui fournit un espace cloud au fournisseur de données pour qu'il puisse y faire l'architecture et la liaison des données.
- *Modèle d'activation* : utilise la virtualisation du cloud pour rassembler les données tout en les conservant dans les limites de l'environnement cloud de l'organisation donatrice.



Mécanisme – Aspects juridiques et gouvernance

M4. Existe-t-il des accords de partage de données entre toutes les parties nécessaires?

Pourquoi c'est important

Les accords de partage de données clarifient le contrôle des données à chaque étape de l'initiative. Cette clarté est importante pour générer la confiance entre les participants. Le statut juridique des accords est nécessaire pour garantir que les participants puissent être tenus pour responsables.

À quoi cela pourrait ressembler

Il existe des accords de partage de données qui définissent clairement et lient légalement les responsabilités et les droits à chaque étape, pour chaque acteur. À titre d'exemple, le Centre de soutien aux espaces de données de l'Union européenne fournit un aperçu du cadre contractuel (Data Spaces Support Centre, 2024) et un résumé de ce que chaque accord contient et est censé faire. De même, ce dépôt fournit des exemples d'accords de partage de données complets conclus par des organisations gouvernementales.

Outils disponibles

D'autres ressources pour soutenir le développement d'accords de partage de données incluent [la conception d'accords de partage de données : une liste de contrôle](#) (Yates et al., 2018) et [une base de données d'accords de partage de données publiés par des organisations gouvernementales](#) (C4DC, n.d).

Recommandation des projets pilotes de feuille de route

Il est plus facile de déterminer quels accords de partage de données seront nécessaires une fois que le contrôle des données le long de la chaîne de publication a été convenu entre les participants (la question 11 a reçu une réponse). Une fois cela convenu, la cartographie des flux de données, des points de décision et des responsabilités de chaque acteur peut aider à identifier où des accords juridiques sont nécessaires.

Mécanisme – Technologie

M5. Existe-t-il une infrastructure de données qui supporte ce mécanisme de partage de données?

Pourquoi c'est important

Un partage de données efficace repose sur une infrastructure de données adaptée aux types de cas d'usage(s) que l'initiative soutient et à l'infrastructure de données existante avec laquelle elle est intégrée.



À quoi cela pourrait ressembler

La conception de l'infrastructure de données pour soutenir une initiative de partage de données comprendra :

- **Intégration des données** : les sources de données et la manière dont elles sont ingérées dans l'installation de stockage des données, y compris les transformations qui doivent être effectuées et les normes qu'elles suivent.
- **Stockage des données** : où les données sont stockées.
- **Découverte des données** : comment les utilisateurs découvrent quelles données sont disponibles et les informations sur ces données.
- **Accès aux données** : comment l'accès des utilisateurs est géré et activé.
- **Services de données** : d'autres services que l'initiative fournit aux utilisateurs de données, tels que des visualisations, des analyses et des applications.

Des exemples de chacune de ces composantes sont explorés plus en détail dans ce Tableau des composantes de l'infrastructure de données, que l'on trouve dans la section [Ressources supplémentaires](#).

Comment y arriver

Pour concevoir l'infrastructure, l'équipe peut travailler avec une équipe de données interfonctionnelle (par exemple, incluant des ingénieurs, des propriétaires, des utilisateurs et du personnel de conformité) afin de comprendre l'infrastructure de données actuelle et élaborer des conceptions futures potentielles. Des illustrations de conceptions d'infrastructures de données qui supportent différents mécanismes de partage de données sont fournies à titre d'exemples dans la section [Ressources supplémentaires](#).

Qu'est-ce qui motive les décisions de conception dans l'infrastructure de données?

1. Infrastructure de données actuelle

- Elle s'intègre aux systèmes existants.

2. Les cas d'usage

- Elle permet aux utilisateurs approuvés d'accéder aux bonnes données grâce à des contrôles d'accès appropriés.
- Elle soutient la gouvernance nécessaire pour les cas d'usage prévus, y compris les mesures de sécurité, la visibilité des flux de données et la facilitation du rôle des éventuels intermédiaires entre les contrôleurs de données et les utilisateurs.
- Elle permet le partage des données à l'échelle et à la fréquence requises par les utilisateurs.



M6. Est-il possible de partager les données de manière sûre et sécurisée?*[Technologie]*

Pourquoi c'est important

Des techniques de préservation de la confidentialité (par exemple, celles établies dans le projet⁵PETs+ du GPAI) et des mesures de cybersécurité⁶ appropriées sont nécessaires pour que les participants puissent avoir confiance dans la sécurité et la robustesse de l'initiative. Des mesures insuffisantes pourraient entraîner des fuites de données ou des violations de la confidentialité qui causeraient un préjudice aux personnes concernées, aux entreprises et aux organisations, ou aux intérêts de la sécurité nationale.

Le ou les cadres juridiques utilisés pour permettre le partage de données au sein de l'initiative peuvent inclure des exigences de confidentialité et de sécurité pour que le partage soit considéré comme légal. L'équipe doit consulter les directives législatives et réglementaires pour s'assurer que toutes les exigences sont respectées.

À quoi cela pourrait ressembler

Il existe des mesures de sécurité des données appropriées pour protéger les données contre les menaces pendant le stockage et la transmission. L'équipe a consulté des responsables gouvernementaux chargés de maintenir les mesures de sécurité des données pour déterminer les besoins uniques du mécanisme et élaborer un plan pour les mettre en place. Ceux-ci incluent des pare-feu, des systèmes de détection d'intrusion et des protocoles de transmission de données sécurisés pour protéger les données contre les menaces de sécurité.

Des techniques de gouvernance technique sont utilisées, le cas échéant, pour préserver la confidentialité des personnes concernées et d'autres intérêts de confidentialité, tels que la sécurité nationale ou les intérêts commerciaux sensibles.

Comment y arriver

En fonction des données partagées, des méthodes appropriées pour garantir la confidentialité des données doivent être identifiées avec des collègues ou des partenaires ayant des compétences techniques.

Méthodes techniques pour préserver la confidentialité des données

Les technologies de préservation de la confidentialité (TPC) sont un ensemble d'outils importants qui permettent ou facilitent le partage sécurisé de données (potentiellement sensibles).

Elles garantissent que la confidentialité des personnes concernées est protégée tout en permettant le partage, le stockage et l'utilisation de ces données ; cette assurance sert à instaurer la confiance entre les personnes concernées, les contrôleurs et les utilisateurs. Les TPC peuvent être particulièrement utiles dans le partage de données gouvernementales pour l'IA, car elles peuvent permettre le partage et l'analyse à grande échelle de données sensibles, par exemple des données au niveau individuel. La pertinence et l'application des différentes

⁵ [GPAI 2023. Overcoming Data Barriers Trustworthy Privacy-Enhancing Technologies, Report, November 2023, Global Partnership on AI.](#)

⁶ [NIST Cybersecurity Framework \(CSF\) 2.0](#)



TPC sont explorées plus en détail dans la ressource TPC que l'on trouve dans la section [Ressources complémentaires](#).

M7. Des pratiques de gestion et de gouvernance des données ont-elles été établies pour garantir la qualité continue des données tout au long de l'initiative? [Technologie]

Pourquoi c'est important

Avant de partager les données, elles doivent être traitées afin d'être dans le format requis pour l'utilisation, et les exigences de formatage dépendent des besoins des utilisateurs de données. L'utilisation de normes communes et existantes peut être importante pour la mise à l'échelle de l'initiative.

À quoi cela pourrait ressembler

Il existe un accord sur les normes de données et le formatage qui seront utilisés au sein de l'initiative et la responsabilité de respecter ces exigences entre les acteurs de l'initiative.

Recommandation des pilotes de la feuille de route

De nombreuses initiatives de partage de données impliquent l'intégration de données actuellement stockées dans différents systèmes. Dans de nombreux cas, les normes de données actuelles appliquées sont différentes d'une organisation à l'autre et d'un système à l'autre, ce qui crée des défis pour l'intégration des données. Par conséquent, les équipes doivent établir des processus et des normes de qualité et d'intégrité des données qui définissent des formats et des schémas de données clairs pour assurer la cohérence et faciliter l'intégration. La mise en œuvre de ces processus doit être négociée entre les équipes, en fonction de leur capacité et de leurs compétences.

Les processus d'intégration des données comprendront des vérifications avant le déplacement des données entre les systèmes, y compris les processus de profilage, de nettoyage et de validation. Ces étapes peuvent être automatisées, semi-automatisées ou effectuées manuellement. Par exemple, les vérifications peuvent être mises en œuvre à l'aide d'outils de validation et de nettoyage des données qui aident à garantir la précision et la cohérence en identifiant et en corrigeant les erreurs, les doublons et les incohérences.

L'intégration des données est également soutenue par des normes et des référentiels de métadonnées gérés de manière centralisée pour aider à la compréhension de la lignée des données, de leurs usages et de leur qualité. Cependant, dans les cas où les données proviennent de sources diverses avec des normes variées, les équipes peuvent adopter une approche centralisée après la collecte des données des sources. Il existe différentes normes *open source* qui peuvent être utilisées pour y parvenir en décrivant la structure et le contenu des ensembles de données.



Mécanisme – Éthique

M8. Le niveau d'ouverture est-il proportionnel au niveau de sensibilité des données?

Le niveau d'ouverture approprié des données est déterminé par leur niveau de sensibilité. Le mécanisme doit s'assurer qu'il rend les données disponibles uniquement aux parties appropriées et utilise des mécanismes d'accès sécurisés et une infrastructure de données de soutien. Cependant, le mécanisme doit en même temps reconnaître que les données gouvernementales ont été payées par le public et doivent donc être rendues aussi ouvertes que possible. L'ouverture des données au public est également étroitement liée aux lois sur l'accès à l'information, et parfois incluse dans celles-ci, qui peuvent comporter des obligations de transparence proactive (Open Data Charter, 2023). Le niveau de sensibilité peut être déterminé par des préoccupations de confidentialité, des intérêts de sécurité nationale ou des droits de tiers sur les données.

À quoi cela pourrait ressembler

L'ouverture des données est priorisée pour permettre un accès aussi large que possible aux données. Ceci afin de promouvoir un accès équitable et d'éviter une distribution inégale des bénéfices de manière inutile. De plus, les mécanismes de partage de données plus ouverts ont tendance à nécessiter des structures de gouvernance plus simples, qui peuvent être plus faciles à gérer et à financer. Cependant, lorsque le ou les cas d'usage soutenus par l'initiative impliquent des données très sensibles, l'ouverture des données doit être limitée et un mécanisme approprié peut être choisi qui prend en charge un accès restreint.



Étape de décision : Structure de gouvernance

La structure de gouvernance définit qui sera responsable de la stratégie globale, de la supervision de l'initiative, et comment cela sera réalisé.

Pourquoi c'est important :

La structure de gouvernance vise à maintenir la confiance tout au long de la vie de l'initiative, tant parmi les participants que le grand public. La structure de gouvernance y parvient en promouvant :

- **L'inclusivité** : en intégrant tous les participants et les groupes affectés et en partageant les bénéfices de l'initiative.
- **la responsabilité** : en assurant une surveillance indépendante.
- **la transparence** : en créant de la visibilité et en s'engageant de manière proactive avec les groupes affectés.
- **la robustesse** : en mettant en place une surveillance et une maintenance continues de l'infrastructure.
- **la gestion des risques** : en établissant des pratiques d'évaluation et d'escalade.

Comment utiliser cette section de la liste de contrôle :

Cette partie de la liste de contrôle peut être complétée pour évaluer une structure de gouvernance existante pour l'initiative de partage de données de l'équipe ou pour guider l'équipe dans le développement d'une telle structure.

Structure de gouvernance – Stratégie et durabilité

G1. La responsabilité de la définition, du suivi et de l'orientation de la stratégie globale de l'initiative est-elle attribuée?

Pourquoi c'est important

Pour soutenir le succès de l'initiative, la responsabilité de la prise de décisions sur son fonctionnement et de la supervision de sa mise en œuvre doit être attribuée. Cela inclut la définition des objectifs de l'initiative, l'élaboration et l'application des règles de prise de décision sur l'accès et l'utilisation des données, et le développement d'un modèle financier durable.

À quoi cela pourrait ressembler

En fonction de l'échelle de l'initiative et des données partagées, ces responsabilités peuvent être confiées à :

- une seule personne ;
- un ou plusieurs groupes sous forme de conseils d'administration ou de comités ; ou
- un tiers de confiance.



Exemple d'établissement de la responsabilité de la stratégie et de la supervision

Initiative (Pays)	Responsabilité de la stratégie et de la supervision
Espace de données sur la mobilité (Allemagne)	L'Espace de données sur la mobilité (MDS) de l'Allemagne, qui facilite le partage de données entre les acteurs privés et publics du secteur de la mobilité, est géré par une entreprise à but non lucratif, la DRM Datenraum Mobilität GmbH. L'entreprise à but non lucratif est gouvernée par trois groupes (Mobility Dataspace, n.d.) : • Un conseil de surveillance élu par les actionnaires, qui surveille et conseille sur la gestion du MDS. • Un conseil consultatif composé d'experts indépendants qui conseillent sur l'orientation stratégique et les nouvelles utilisations du MDS. • Des actionnaires, y compris des organisations des secteurs privé et public.
INSIGHT (Royaume-Uni)	INSIGHT est un fonds de données géré par le Service national de santé (NHS) du Royaume-Uni pour permettre le partage responsable de données de santé oculaire très sensibles afin de faciliter la recherche. La gouvernance du fonds de données est coordonnée par le NHS, tandis qu'un conseil consultatif, le Data Trust Advisory Board (DataTAB), fournit un examen indépendant et des recommandations consultatives sur toutes les demandes d'utilisation des données reçues par INSIGHT (INSIGHT, n.d.).

G2. Un budget est-il garanti pour cette initiative de partage de données, ou y a-t-il des options pour son financement?

Pourquoi c'est important

L'initiative a besoin d'un financement pour un projet pilote et d'options pour garantir un financement à long terme afin d'assurer sa réalisation. Sans ce financement, il ne sera pas possible de garder les bonnes personnes impliquées, de se procurer la bonne technologie pour l'infrastructure de données ou de mener des consultations avec les parties prenantes.

À quoi cela pourrait ressembler

Le financement peut provenir du gouvernement, de banques de développement ou être de source privée. Il peut donc se présenter sous la forme d'une contrainte budgétaire imposée par des bailleurs de fonds du gouvernement central, de subventions ou d'un contrat privé. Bien qu'il n'y ait pas de budget à long terme garanti, les équipes devront avoir des options pour obtenir un budget pour une démonstration pilote de l'initiative. Les équipes devraient également commencer à envisager leurs options pour assurer la viabilité et l'évolutivité du projet à plus long terme (GPAI, 2022b).



Structure de gouvernance – Communauté et Parties prenantes

G3. Existe-t-il des mécanismes de communication des représentants de tous les acteurs affectés par l'initiative vers les personnes responsables de sa supervision?

Pourquoi c'est important

Les personnes responsables de la stratégie et de la supervision peuvent ne pas inclure de représentants de tous les groupes affectés par l'initiative. La représentation et l'inclusion de ces groupes affectés sont importantes pour garantir que l'initiative œuvre au service de l'intérêt public.

À quoi cela pourrait ressembler

Établir des mécanismes de communication signifie créer des canaux pour que l'information circule des groupes affectés vers les responsables de la stratégie et de la supervision. Cela peut signifier avoir des représentants de groupes au sein d'un conseil, organiser des réunions régulières ou utiliser des méthodes de collecte de données telles que des sondages. Il est important de prendre en compte les groupes suivants :

- les dirigeants du gouvernement ou de l'agence ;
- les équipes centrales à l'initiative ;
- les contributeurs de données ;
- les utilisateurs de données ;
- toute personne concernée par les données ; et
- les instances de réglementation.

G4. La transparence est-elle maintenue tout au long du développement et de la mise en œuvre de l'initiative? *[Communauté et Parties Prenantes]*

Pourquoi c'est important

La transparence concernant l'initiative est nécessaire pour la responsabilité. Elle est nécessaire pour que les participants puissent s'engager et contester sa gestion, pour que le public puisse exprimer ses préoccupations et pour que les autorités indépendantes puissent faire respecter les réglementations pertinentes. En plus de faciliter la responsabilité en cas de préoccupations, la transparence génère également la confiance et aide l'équipe à sensibiliser aux bénéfices de l'initiative.

À quoi cela pourrait ressembler

Atteindre la transparence nécessite différentes actions pour différents acteurs et cas d'usage. Les actions doivent permettre aux acteurs d'être suffisamment bien informés pour faire confiance à l'initiative, et ce qui est requis peut être déterminé en interagissant avec les différents groupes d'acteurs.

Pour certains acteurs, cela signifiera de partager les progrès et de communiquer les décisions. Selon le contexte, des exemples incluent la publication des procès-verbaux des réunions du conseil de gouvernance ou l'organisation de présentation et d'échanges pour les participants.

Pour les personnes concernées par les données en particulier, la transparence exige qu'elles aient une visibilité sur qui détient leurs données, dans quel but et qu'elles soient informées des éventuels transferts de données. Cela pourrait être réalisé grâce à des systèmes de gestion du consentement électronique (République d'Estonie, 2024), ou facilité par la législation sur l'accès à l'information, et mis en œuvre par le biais de demandes d'information.



G5. Existe-t-il un plan pour permettre à un large éventail d'acteurs d'accéder aux bénéfices de l'initiative?

[Communauté et Parties Prenantes]

Pourquoi c'est important

Pour que l'initiative maximise sa contribution à la résolution du ou des problèmes d'intérêt public qu'elle cible, les équipes peuvent réfléchir à la manière dont les bénéfices peuvent être largement répartis entre les acteurs. Les bénéfices peuvent inclure l'accès aux données elles-mêmes, les gains monétaires tirés des données, des systèmes d'IA, et des produits, services ou recherches développés à l'aide de systèmes d'IA. Il est important d'interpréter les bénéfices au sens large afin qu'ils puissent être répartis équitablement entre les groupes affectés et le grand public.

À quoi ça pourrait ressembler

L'équipe a réfléchi à la manière de rendre les bénéfices accessibles au plus grand nombre d'acteurs possible. La façon dont les bénéfices peuvent être distribués dépendra des données partagées et des résultats de l'initiative. Cependant, quelques exemples d'approches liées à différents types d'initiatives sont détaillés ci-dessous.

Exemples d'augmentation de l'accès aux bénéfices

Approche	Exemples
Informer les utilisateurs de la disponibilité des données	Le programme Copernicus de l'UE fournit des bulletins d'information sur la publication de nouveaux ensembles de données, auxquels les utilisateurs peuvent s'abonner pour recevoir des alertes (Copernicus, n.d.). De nombreux portails de données ouvertes et initiatives de partage de données avec des membres partagent les mises à jour de la même manière.
Publier les résultats du partage de données	Les résultats peuvent inclure de la recherche universitaire ou des études gouvernementales. Aux États-Unis, les publications et les recherches financées par les contribuables doivent être rendues immédiatement accessibles au public (The White House, 2022). D'autres types de résultats peuvent inclure des informations qui éclairent l'élaboration de politiques et qui sont destinées à un usage interne au gouvernement. Les équipes pourraient envisager de visualiser l'analyse des données et de partager les résultats au sein du gouvernement, par exemple avec des tableaux de bord.
Concevoir des services inclusifs	Si les données sont utilisées dans des systèmes d'IA pour la prestation de services gouvernementaux, des techniques de conception centrées sur l'humain et sur l'utilisateur peuvent être utilisées pour garantir que les systèmes sont intégrés d'une manière qui répond aux besoins de tous les groupes d'utilisateurs (Chen et al, 2024; Gov.UK, 2024).



Structure de gouvernance – Aspects juridiques et gouvernance

G6. L'initiative recueille-t-elle le consentement des personnes concernées par les données et leur permet-elle de le retirer à tout moment?

Pourquoi c'est important

Le consentement est nécessaire dans presque toutes les circonstances où des données personnelles sont partagées (GPAI, 2023). Ceci est important pour respecter les droits des personnes concernées et, dans la plupart des juridictions, pour se conformer à la législation sur la protection des données et la confidentialité.

À quoi cela pourrait ressembler

Les exigences spécifiques pour la collecte du consentement, et ce qui est suffisant pour le consentement, diffèrent selon les juridictions. Les exigences seront incluses dans la réglementation sur la protection des données et la confidentialité de la juridiction, et dans toute directive d'accompagnement, ainsi que dans toute réglementation sectorielle spécifique, par exemple, relative aux données de santé. L'équipe doit s'engager avec les régulateurs de la protection des données pour comprendre comment cela s'applique dans le cas spécifique de l'initiative.

Le consentement peut être demandé au moment de la collecte des données, ou les personnes concernées peuvent ensuite se voir offrir la possibilité de s'inscrire ("opt-in") ou de se désinscrire ("opt-out") d'une initiative de partage de données. Les équipes doivent s'assurer que le consentement recueilli couvre la portée et l'objectif du traitement des données impliqué dans le ou les cas d'usage de l'initiative. Si l'équipe utilise une approche de "désinscription" pour établir le consentement, cela nécessite des communications actives aux personnes concernées au sujet de l'initiative et de l'option de se désinscrire.

G7. Existe-t-il un mécanisme par lequel les personnes concernées par les données peuvent demander réparation pour des violations de leurs droits?

Pourquoi c'est important

Même si les risques doivent être atténués et qu'aucun risque inacceptable ne doit être pris par l'initiative, il doit y avoir des mécanismes permettant aux personnes concernées par les données de soulever des préoccupations concernant leurs droits et de demander réparation si leurs droits sont violés. Cette forme de responsabilité est importante pour garantir que l'initiative reste dans les limites juridiques et éthiques du partage de données et de l'utilisation et du développement de l'IA. Il est également important pour maintenir la confiance dans l'initiative : les individus doivent savoir qu'ils ont un recours au cas où quelque chose tournerait mal.

À quoi cela pourrait ressembler

Les mécanismes qui permettent aux personnes concernées par les données de demander réparation peuvent être activés par la législation et facilités par l'initiative, par exemple en publiant de manière proactive des informations sur la manière de faire part de ses préoccupations. Les exemples incluent la législation de l'UE sur la protection des données qui restreint la prise de décision automatisée en ce qui concerne certaines décisions affectant les personnes concernées (Commission européenne, n.d.).



G8. Existe-t-il un canal pour que les agences de réglementation pertinentes puissent suivre l'initiative et donner des conseils sur son développement?

Pourquoi c'est important

L'inclusion des agences de réglementation tout au long du développement et de la mise en œuvre de l'initiative est un facteur critique de la surveillance indépendante (GPAI, 2023).

À quoi cela pourrait ressembler

L'équipe a fourni une divulgation complète aux autorités de surveillance pour qu'elles examinent la conception de l'initiative et reçoivent des conseils sur sa mise en œuvre, même lorsque les cadres réglementaires n'exigent pas de telles divulgations. Des processus sont mis en place pour une communication et un examen continus de l'initiative avec les régulateurs tout au long de sa vie.

Recommandation des projets pilotes de feuille de route

Avoir des représentants de l'agence ou des agences de réglementation pertinentes dans la salle lors des réunions clés peut accélérer la mise en œuvre du projet. C'est parce qu'ils ont de l'expérience dans l'application des cadres juridiques et réglementaires pertinents. Cela peut aider à clarifier les possibilités et les limites de l'initiative dès le début.

G9. Existe-t-il des processus en place pour surveiller la conformité aux accords de partage de données?

Pourquoi c'est important

Ceci est important pour gérer les risques qui peuvent découler du non-respect par les parties des conditions d'utilisation des données établies dans les accords de partage de données.

À quoi cela pourrait ressembler

Il existe plusieurs voies pour surveiller la conformité.

Exemples de mécanismes de surveillance de la conformité aux accords de partage de données

Type de mécanisme	Mécanismes de surveillance
Processus opérationnels	• Détruire les données et désactiver l'accès aux environnements de partage de données à la fin de la période de licence. • Mettre en place des examens réguliers des accords pour s'assurer qu'ils continuent de répondre aux besoins des parties impliquées.
Caractéristiques techniques de l'environnement de partage de données	• Enregistrer et journaliser automatiquement l'accès et le traitement des données par des tiers.



Type de mécanisme	Mécanismes de surveillance
	• Empêcher le téléchargement de données sur des appareils personnels ou le copier-coller de données en dehors d'un environnement de partage de données.
Législation	• Politiques de contrôle de la divulgation pour évaluer les résultats de l'utilisation des données.
Application réglementaire	• L'agence de réglementation a le pouvoir de mener des audits des processeurs autorisés.
Dénonciation	• Créer des canaux pour que les acteurs puissent faire part de leurs préoccupations concernant une éventuelle non-conformité.

Structure de gouvernance – Technologie

G10. Existe-t-il des processus en place pour surveiller l'infrastructure et maintenir un partage de données sûr et sécurisé pendant toute la durée de l'initiative?

Pourquoi c'est important

Pour que les participants soient disposés à partager des données au sein de l'initiative, il faut une infrastructure de données robuste et fiable. Cela nécessite une surveillance et une maintenance continues.

À quoi cela pourrait ressembler

L'infrastructure de données comprend des serveurs pour le stockage des données, le traitement et l'accès pour les utilisateurs de données. La surveillance continue peut impliquer l'utilisation d'outils de surveillance des performances et la réalisation d'analyses de sécurité. Les processus que le gouvernement prend en charge dépendront du type de serveur utilisé et de son emplacement.

Des processus sont mis en place pour garantir que l'accès est donné aux utilisateurs autorisés et que l'accès est limité aux utilisateurs autorisés. Les processus dépendront du mécanisme d'accès, par exemple, les API nécessiteront des contrôles réguliers de leur état de fonctionnement.

G11. Le gouvernement est-il équipé pour mener un processus d'approvisionnement équitable qui évalue les systèmes d'IA afin de s'assurer qu'ils répondent aux exigences de l'initiative?

Pourquoi c'est important

De nombreuses initiatives de partage de données impliqueront l'approvisionnement de partenaires externes pour utiliser les données gouvernementales afin de fournir des produits ou services basés sur l'IA. Le gouvernement doit être en mesure de mener un processus d'approvisionnement équitable qui évalue efficacement les systèmes d'IA pour s'assurer qu'ils



peuvent répondre aux exigences de l'initiative. Cela inclut les exigences fonctionnelles et de performance ainsi que les exigences éthiques et de sécurité.

À quoi cela pourrait ressembler

Pour que le processus d'approvisionnement soit équitable, il devrait y avoir des processus d'approvisionnement standardisés qui sont visibles et interprétables par le public. Il peut y avoir des dispositions pour encourager les petites entreprises à soumissionner.

Pour l'évaluation efficace des systèmes d'IA, il existe des processus d'évaluation appropriés, que les systèmes d'IA soient déployés à l'intérieur ou à l'extérieur du gouvernement et pour les tâches pour lesquelles les systèmes sont utilisés. L'utilisation d'un système à l'intérieur du gouvernement nécessitera une plus grande transparence de la part du fournisseur du système. Cela pourrait inclure l'utilisation de fiches de modèle ou d'autres techniques de rapport de modèle pendant l'approvisionnement.

Une évaluation des fournisseurs au gouvernement pourrait être menée pour s'assurer qu'ils sont conformes aux cadres éthiques et aux normes de service du gouvernement ou internationalement reconnus.

La réalisation des évaluations exige que les responsables de l'approvisionnement aient accès à une expertise en technologie de l'IA et à ses risques. Ils peuvent avoir accès à d'autres fonctionnaires ayant une expertise technique pertinente, avoir reçu une formation spécialisée ou avoir des directives à suivre.

Outils Disponibles

Des ressources supplémentaires qui peuvent aider les équipes à développer des processus d'évaluation des systèmes d'IA dans le cadre de leur processus d'approvisionnement incluent les [Principes FAIR](#) (Go Fair, n.d.), [les Fiches de modèle pour la création de rapports](#) (Mitchell et al., 2019), [AI Procurement in a Box](#) (World Economic Forum, 2020), [AI Toolkit Procurement Guide](#) (Tony Blair Institute for Global Change, 2022).

Structure de gouvernance – Éthique

G12. Existe-t-il des processus pour surveiller et répondre aux risques pour les personnes concernées par les données et les autres groupes affectés?

Il est important d'établir des processus de surveillance et de gestion des risques tout au long de la mise en œuvre de l'initiative. Ces processus peuvent être ensuite examinés par les responsables de la stratégie et de la supervision. L'objectif final est de s'assurer que les risques pour les droits des personnes concernées par les données et des autres groupes sont prévenus et ainsi garantir que l'initiative atteigne ses bénéfices d'intérêt public déclarés.

À quoi cela pourrait ressembler

Les processus seront déterminés par les risques uniques identifiés dans un registre des risques, présenté à la [question 9](#), pour l'initiative. Cependant, l'équipe se sera mise d'accord sur la manière dont les risques seront surveillés et aura élaboré un processus d'escalade pour savoir comment réagir si des risques surviennent.



La surveillance de l'évolution des risques peut être facilitée par des conditions établies dans les accords de partage de données. Par exemple, les accords peuvent exiger la publication ou la transmission de rapports sur les résultats de l'utilisation des données à l'initiative.

Les collaborateurs sont susceptibles de faire face à un compromis entre l'efficacité opérationnelle et la surveillance. Répondre à ce compromis nécessitera une hiérarchisation des risques afin que la surveillance soit construite autour de ceux considérés comme critiques.

G13. L'initiative exige-t-elle une surveillance humaine pour toute prise de décision algorithmique?

Pourquoi c'est important

Cette question aborde le besoin de surveillance humaine quant à la manière dont les résultats des algorithmes alimentent la prise de décision gouvernementale, que ce soit dans la prestation de services publics, les opérations en arrière-plan, ou sous la forme d'analyses de données qui éclairent les décisions politiques.

Une surveillance humaine des résultats des algorithmes d'IA et de la manière dont ces résultats sont utilisés dans les décisions est nécessaire pour corriger les erreurs de la machine et prévenir les effets indésirables sur les personnes concernées par les données, tels que l'exclusion d'un service public.

De plus, déléguer l'autorité sur une décision qui affecte un individu à un système d'IA peut être inacceptable dans des contextes où des cadres de justice administrative existent.

À quoi cela pourrait ressembler

Les responsables de la stratégie et de la supervision précisent comment les résultats algorithmiques seront utilisés dans la prise de décision afin de maintenir la surveillance humaine. Les données peuvent être partagées pour une application spécifique de l'IA au sein du gouvernement ; dans ce cas, la spécification de la manière d'utiliser les résultats peut être adaptée au cas d'usage. Par exemple, cela peut signifier fournir des directives sur la manière dont les fonctionnaires doivent interpréter les résultats des systèmes d'IA, y compris leurs marges d'erreur et leurs limites.

Dans les cas où des données sont partagées pour être utilisées dans le développement de l'IA de manière plus large, par exemple par des chercheurs et des organisations en dehors du gouvernement, la spécification peut prendre la forme de directives que les utilisateurs acceptent avant d'accéder aux données.

Outils disponibles

Des ressources supplémentaires qui peuvent aider les équipes à déterminer comment maintenir une surveillance humaine dans leur initiative incluent [Vers une surveillance significative des systèmes de prise de décision automatisée](#) (Digital Future Society, 2022).

Voir également le rapport du GPAI, [le rapport de projet sur la transparence algorithmique dans le secteur public](#) (2022).

Études de cas pilotes

L'objectif de la deuxième phase du projet était de concevoir un ensemble de documents d'orientation, une liste de contrôle et une feuille de route, que les organisations gouvernementales pourraient utiliser pour faciliter le partage de données pour une IA axée sur l'intérêt public. Pour ce faire, nous avons identifié la nécessité de tester toute orientation que nous développerons auprès des personnes impliquées sur le terrain dans des projets de partage de données pour l'IA.

Nous avons établi trois partenariats pilotes pour effectuer les tests des documents. Les partenariats ont donné aux pays pilotes l'occasion de travailler sur des projets de partage de données existants et de les faire progresser. Ils leur ont également permis d'accéder à des conseils d'experts sur la gouvernance et l'infrastructure technique pour le partage de données, les aidant à développer leurs capacités internes.

Parallèlement, ils nous ont permis de tester la feuille de route auprès de ceux qui, idéalement, l'utiliseraient et d'identifier les domaines d'amélioration. Nous avons pu tester à la fois le contenu et la structure de la feuille de route, en précisant si les informations que nous fournissions étaient réellement utiles et si elles étaient transmises de la manière la plus appropriée.

Comment nous avons choisi les partenaires pilotes

Pour nous assurer que ce travail avait une large portée et une pertinence, par exemple sur les plans économique, géographique, gouvernemental et sectoriel, nous avons élaboré un ensemble de critères pour nous assurer que nous avions une répartition représentative parmi les partenaires pilotes. Nous avons lancé un appel à manifestation d'intérêt et avons reçu plus de 20 réponses d'agences gouvernementales fédérales, nationales et municipales de plusieurs pays et avec un éventail d'intérêts. Nous avons noté les réponses des pays, en examinant la capacité qu'ils avaient de s'engager de manière significative à toutes les étapes de la feuille de route provisoire que nous avons élaborée. De plus amples informations sur ce processus sont présentées dans la section [Méthodologie](#), Annexe 2.

À la suite de ce processus d'engagement, nous avons choisi trois partenaires pilotes :

1. le gouvernement de l'État de Jigawa au Nigeria;
2. le Bureau de la transformation numérique de la Türkiye; et
3. l'Agence pour le gouvernement électronique et la société de l'information (AGESIC) en Uruguay.

Ces partenaires pilotes représentaient un éventail de niveaux de gouvernement national et fédéral ainsi qu'un intérêt pour un large éventail de secteurs et de cas d'usage. Chaque engagement a suivi à peu près la même structure : des points de contact réguliers avec les partenaires; une sélection d'ateliers sur des sections de la feuille de route; et des sessions individuelles avec des experts en données et en politiques d'Oxford Insights.

Le reste de cette section donnera une description de chaque pilote individuel, expliquant :

- **le contexte spécifique du pays** : motivations et point de départ;
- **le résultat de la mise en œuvre de la feuille de route** : le problème que le cas d'usage tente de résoudre, le type de mécanismes discutés, la gouvernance impliquée; et
- **les activités pilotes que nous avons menées avec les pays** et les parties prenantes impliquées.

Enseignements tirés de la mise en œuvre de la feuille de route pilote

En mettant en œuvre la feuille de route avec nos partenaires pilotes, nous avons identifié un certain nombre d'opportunités et de défis clés pour les équipes qui entreprennent des initiatives de partage de données pour l'IA, ainsi qu'une meilleure compréhension de la manière dont la feuille de route peut aider les équipes tout au long de leurs initiatives.

Opportunités :

La feuille de route est une opportunité pour les équipes d'accélérer les initiatives numériques existantes en fournissant un guide clair pour avancer et pour évaluer de manière systématique les décisions qui avaient déjà été prises dans l'initiative, si elle a déjà commencé. L'Uruguay et le Jigawa ont tous deux commencé avec des secteurs dans lesquels ils voulaient développer des initiatives. Ils voulaient tous deux commencer par **circonscrire** leur approche en partant d'un secteur pour aboutir à un cas d'usage, puis **déterminer** un mécanisme de partage de données en fonction de cela. La Türkiye a commencé avec un mécanisme, et a choisi des cas d'usage en fonction des données qui pouvaient être partagées étant donné l'infrastructure et le cadre juridique qui soutenaient ce mécanisme. Nous avons mis à jour la feuille de route pour s'adapter aux différents points d'entrée des partenaires et aux différentes séquences de prise de décision préférées.

Le partage de données pour les initiatives d'IA nécessite une forte collaboration intergouvernementale, et parfois l'attribution de nouvelles responsabilités pour les parties prenantes. L'évaluation initiale dans la feuille de route peut aider à sensibiliser aux opportunités de partage de données pour l'IA avec des partenaires potentiels intergouvernementaux. Il est utile de maintenir ce niveau d'engagement tout au long de l'initiative afin que les rôles dans l'initiative, y compris la préparation des données et la gestion de l'accès, puissent être convenus plus facilement.

Il existe de nombreux modèles émergents de partage de données pour l'IA et ceux-ci peuvent être adaptés aux besoins de l'initiative et au contexte gouvernemental. L'expérimentation au sein de modèles de partage de données à petite et grande échelle à travers les gouvernements peut être utile comme exemples de la manière dont les équipes organisent les rôles des participants et les procédures au sein de leurs propres initiatives. Cependant, il est important de les adapter au contexte spécifique, y compris l'infrastructure de données existante, les organisations qui jouissent d'une grande confiance dans ce domaine et les canaux de communication existants entre les parties.

Les différentes étapes de la feuille de route peuvent être mises en œuvre simultanément. Par exemple, le fait de décider de partager des données sensibles peut soulever des inquiétudes quant à la capacité de l'équipe à mettre en place les structures de gouvernance techniques et opérationnelles solides requises. Elles peuvent vouloir résoudre ces préoccupations dès le début avant de progresser.

Défis :

La nouveauté des initiatives de partage de données pour l'IA signifie que les équipes ont une expérience limitée dans l'application des cadres juridiques applicables. Même lorsqu'il existe un cadre juridique et réglementaire national clair, l'expérience de l'application de ce cadre au sein de l'équipe, et au cas spécifique d'une nouvelle initiative, peut être limitée. Cela rend d'autant plus important le fait que les équipes demandent directement des conseils aux agences de réglementation, mais souligne également la nécessité pour ces agences de développer des directives applicables et réutilisables sur le partage de données dans leurs juridictions qui peuvent être utilisées comme points de départ pour les équipes.

Les équipes manifestent un vif intérêt pour l'application des techniques de gouvernance technique émergentes, mais le savoir-faire en matière de mise en œuvre est limité. Par exemple, les équipes sont désireuses d'adopter des techniques de préservation de la confidentialité, telles que l'apprentissage fédéré ou les données synthétiques, qui peuvent leur permettre de partager des données sensibles avec une plus grande certitude de sécurité. Cependant, ce sont de nouvelles techniques et les équipes, même si elles en ont connaissance, n'ont généralement pas d'expérience dans leur mise en œuvre. Ce défi appelle au partage des meilleures pratiques des gouvernements qui emploient ces techniques et à de plus grandes opportunités pour les équipes de perfectionner leurs compétences techniques dans les technologies émergentes.

Limites de la feuille de route :

Une limitation clé de notre processus de conception pour les projets pilotes a été que, en raison de contraintes de temps, nous n'avons interagi qu'avec les parties prenantes gouvernementales. Cela a créé une limite quant à la mesure dans laquelle nous pouvions progresser dans la conception des initiatives. Dans tous les cas, nos partenaires pilotes s'engagent dans des consultations externes au cours des phases suivantes de leurs projets.

Une autre limitation de la feuille de route est que, en tant qu'outil destiné à un usage mondial, elle ne peut pas aborder les défis qui résultent de contextes gouvernementaux uniques. En tant que ressource destinée à un usage mondial, elle devait être, à notre connaissance, compatible avec tous les contextes juridiques, institutionnels et techniques. La feuille de route vise à atténuer les inconvénients de cette nécessité en fournissant des exemples pratiques de la manière dont les gouvernements ont réagi aux défis dans leurs propres contextes.

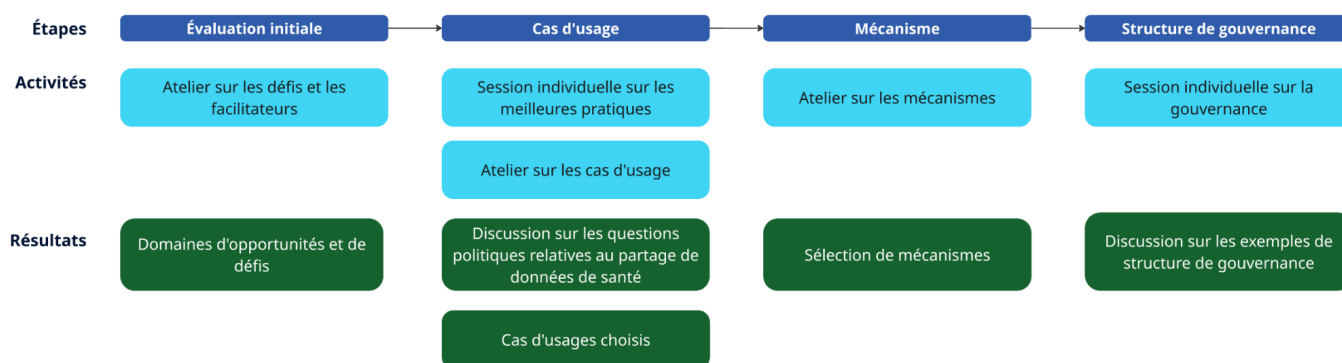
Étude de cas : Uruguay

Motivations et point de départ

L'Agence uruguayenne pour le gouvernement électronique et l'information (AGESIC) a répondu à notre appel à manifestation d'intérêt, exprimant son intérêt pour le partage de données afin de soutenir le développement d'outils d'IA visant à lutter contre le changement climatique, ainsi que pour le partage de données pour la santé numérique. Ils disposaient de plusieurs ensembles de données climatiques accessibles au public et souhaitaient explorer des moyens de partager des données de santé avec des parties prenantes non gouvernementales.

Activités pilotes et résultats

Nous avons organisé trois ateliers avec l'équipe d'AGESIC, ainsi que deux sessions individuelles détaillées avec des experts d'Oxford Insights. Certaines sessions ont également été suivies par l'équipe de santé numérique au sein d'AGESIC et par l'Autorité uruguayenne de protection des données (URCDP).

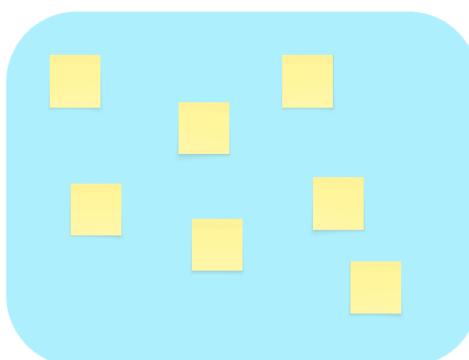


Évaluation Initiale

Nous avons commencé par mener un atelier stratégique basé sur la première section de notre projet de feuille de route pour déterminer la capacité et les priorités stratégiques au sein du gouvernement et d'AGESIC.

1. Stratégie et durabilité

- Y a-t-il des objectifs clairs pour l'initiative?
- Un budget est-il garanti pour l'initiative de partage de données ou y a-t-il des options pour son financement?
- Y a-t-il des options pour garantir un budget à long terme pour la pérennité et l'extension de l'initiative?



En particulier, nous avons examiné les défis et les facilitateurs dans tous les secteurs. Nous avons produit un tableau Miro avec les domaines d'opportunités et de défis dans tous les secteurs. Les opportunités étaient liées aux éléments suivants :

- Le fait d'avoir déjà été approché par des parties intéressées pour obtenir l'accès à des données dans certains secteurs
- Un cadre juridique solide en place pour l'échange de données et les informations personnelles
- L'expérience dans la formation d'accords de partage de données
- Le fait que la plateforme d'interopérabilité du gouvernement permette aux organisations publiques d'échanger des données en toute sécurité

Pendant ce temps, les défis étaient centrés sur :

- Le financement qui dépend des budgets d'autres départements
- Une compréhension limitée des bénéfices par le public
- Une expérience limitée dans l'application des cadres juridiques pertinents dans le contexte de l'utilisation de l'IA
- L'obtention et la gestion du consentement éclairé

Bien que le changement climatique soit également d'intérêt, à la suite de l'atelier sur la stratégie et du travail indépendant sur les sections initiales de bilan et de sélection des secteurs de la feuille de route, AGESIC a finalement décidé de se concentrer sur la santé numérique. Le Ministère Uruguayen de la Santé détient une grande quantité de données provenant de diverses sources, notamment des cliniques et des hôpitaux. Ils étaient conscients de leur valeur et du potentiel qu'elles avaient pour améliorer un certain nombre de processus/modes d'analyse différents, y compris la santé publique et le suivi des maladies. Cependant, ils ne savaient pas comment structurer le programme de partage de données qui permettrait aux données détenues de manière centralisée par le ministère de la Santé d'être distribuées à d'autres parties qui pourraient les utiliser et de quelles manières.

Cas d'usage

Nous avons organisé un atelier où nous avons travaillé ensemble pour cartographier les parties prenantes pertinentes dans le secteur de la santé, les flux de données et affiner les types de problèmes spécifiques qui pourraient être résolus dans ce secteur. Cela nous a permis de nous concentrer sur le partage de données dans les systèmes prédictifs de santé; par exemple, la prédiction des maladies et l'imagerie médicale diagnostique. Deux facteurs principaux ont influencé la décision concernant le cas d'usage. Premièrement, il y avait un besoin pressant d'améliorer les systèmes de prédiction des maladies pour optimiser la planification stratégique et l'allocation des ressources. De plus, le ministère de la Santé détenait des données pertinentes qui pouvaient être exploitées par les développeurs d'IA pour construire ces systèmes. Ce cas d'usage répondait aux exigences de la liste de contrôle : une forte demande en données, une offre de données prêtes pour l'IA et une adéquation avec les priorités stratégiques.



Cette décision concernant le cas d'usage a soulevé un certain nombre de considérations de gouvernance spécifiques au partage de données de santé que l'équipe a souhaité examiner avant d'aller de l'avant. En particulier, des considérations relatives aux éléments suivants :

- Les facteurs affectant la portée de l'utilisation des données par les destinataires, notamment le type de données partagées (données personnelles identifiables (DPI), données anonymisées), qui sont les utilisateurs des données (universitaires, chercheurs, entreprises privées), la manière dont la prise de décision a lieu (fiduciaires de données, membres de coopératives), les contrôles d'accès aux données en place et ce à quoi les fournisseurs de données ont consenti et comment (modèles d'adhésion/de refus)
- L'assurance que les données qui ont été partagées sont gérées en toute sécurité grâce à (1) la conformité avec la Législation Nationale sur la Protection des Données et (2) le respect des Normes de Partage de Données

Mécanisme et gouvernance

Ayant déterminé un cas d'usage, nous avons ensuite parcouru les phases suivantes de la feuille de route : nous avons travaillé à l'identification d'un mécanisme de partage de données qui serait le mieux adapté à ce type de partage de données, en utilisant la feuille de route ainsi que des exemples concrets de partage de données de santé, tels que la coopérative de données suisse MIDATA et l'initiative de gestion des données du NHS britannique, INSIGHT.

Cela nous a amenés à choisir trois mécanismes potentiels de partage de données : les fonds de données (data trusts), les gestions de données (data stewardships) et les biens communs de données (data commons), en nous basant sur la pertinence de ces mécanismes pour des facteurs tels que la sensibilité des données, le rôle des participants dans la prise de décision et le rôle du gouvernement dans l'initiative.

 Características de los mecanismos				
Dimensión	Data Trust	Data Stewardship	Data Cooperative	Data Commons
Decisiones sobre acceso y uso de datos	Centralizado	Centralizado	Descentralizado	Descentralizado
Confianza entre poseedores y usuarios	Bajo	Bajo	Medio	Medio/Alto
Sensibilidad de los datos	Alta	Media/Alta	Media/Alta	Baja
Nivel de involucramiento de los poseedores	Bajo	Medio	Alto	Alto
Capacidad digital de los poseedores	Baja	Bajo/Medio	Alta	Media/Alta
Rol potencial del gobierno	Proveedor de infraestructura, regulador	Coordinador, regulador	Coordinador, facilitador	Proveedor de infraestructura, facilitador



Enfin, nous avons organisé un atelier pour examiner les aspects pratiques de chacun de ces mécanismes et les flux de données dans le secteur de la santé. À ce stade, AGESIC a pris la décision de suspendre les travaux de la feuille de route, car ils estimaient qu'il serait nécessaire de revenir aux parties prenantes et de mener des recherches plus détaillées pour déterminer le mécanisme le plus approprié, avant de progresser.

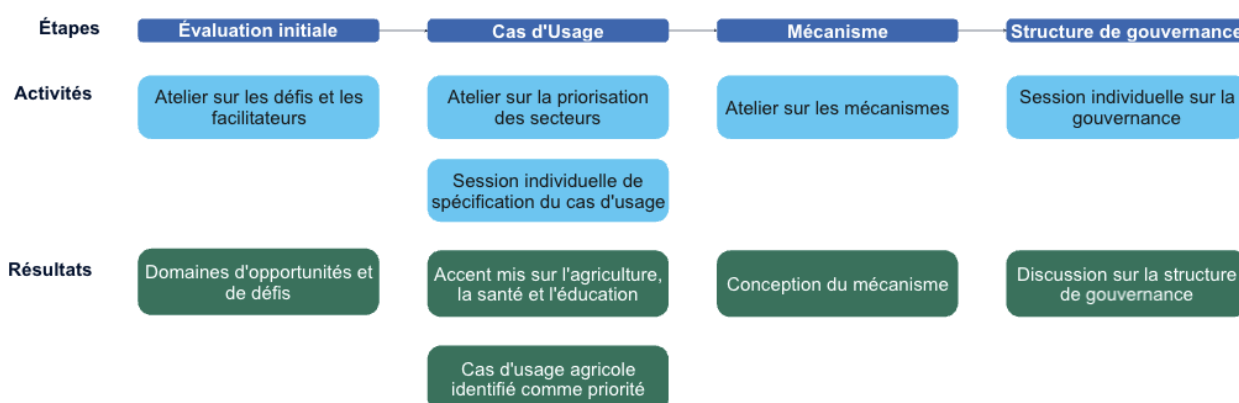
Étude de cas : Jigawa

Motivations et point de départ

Jigawa nous a approchés avec une équipe diversifiée et multidisciplinaire, dont les membres couvrent un certain nombre de secteurs et d'expertises. Au début du projet, l'équipe avait commencé à exploiter les données en collaboration avec un certain nombre d'organisations internationales et avait mis en place certaines pratiques de partage de données pour le partage entre les équipes au sein du gouvernement. Cependant, ils cherchaient à accélérer la numérisation générale au sein du gouvernement de l'État et à utiliser l'IA pour résoudre les défis de développement. Ils ont travaillé avec nous pour aider à identifier les secteurs/cas d'usage les plus prometteurs pour le partage de données afin de développer des outils d'IA, et créer un mécanisme de prise de décision sur mesure au sein du programme.

Activités pilotes et résultats

Nous avons organisé trois ateliers avec l'équipe de Jigawa ainsi que deux sessions individuelles détaillées avec des experts d'Oxford Insights. Certaines sessions ont également été suivies par des représentants des unités de l'agriculture, de la planification de la main-d'œuvre, de la santé et de l'éducation au sein du gouvernement de l'État.



Évaluation Initiale

Nous avons organisé un atelier sur les défis et les facilitateurs pour identifier les défis et les facilitateurs transversaux ainsi que les priorités stratégiques. Ceux-ci ont été résumés dans un document que le Jigawa pourrait utiliser pour éclairer les futures décisions sur la priorisation des secteurs et la spécification des cas d'usage. Les domaines d'opportunités mis en évidence étaient :

- L'équipe a établi des relations entre les équipes et les agences.
- Il y a une adhésion du leadership et le projet bénéficie d'un soutien politique important de la part des décideurs de haut niveau.
- Des données sont collectées qui pourraient être utiles pour répondre aux défis dans un certain nombre de secteurs, y compris la santé, l'éducation et l'agriculture.

- Certaines pratiques de partage de données ad hoc sont en place entre les équipes gouvernementales, généralement effectuées par transfert de fichiers. Pour l'avenir, il y a une ambition de créer un stockage central de données et de faciliter l'accès à travers le gouvernement en utilisant des APIs.
- Il y a un élan pour l'élaboration de politiques visant à soutenir une transition vers un gouvernement plus numérique, et le gouvernement commence à introduire des politiques de gestion des données qui soutiendront la standardisation des données gouvernementales.

Pendant ce temps, les domaines mis en évidence comme des obstacles potentiels étaient :

- Le gouvernement a de nombreuses priorités et ne dispose pas d'un financement disponible pour les soutenir.
- L'accès limité aux compétences en IA signifie que l'équipe aurait besoin de partenaires de mise en œuvre pour construire et implémenter des applications d'IA au sein du gouvernement.
- Il y a encore une sensibilisation limitée au sein de certaines équipes concernant la transition vers un gouvernement numérique et les risques et pratiques de gouvernance qui devront l'accompagner.
- La plupart des opérations et des services gouvernementaux sont encore "analogiques". Cela limite les données disponibles, et il existe une infrastructure de partage de données limitée sur laquelle s'appuyer pour faciliter ce nouveau projet.
- Il existe une autorité de protection des données à l'échelle de l'État responsable de l'application de la loi nigériane sur la protection des données. Cependant, il s'agit d'une nouvelle agence avec une expérience limitée et une relation avec l'équipe doit encore être établie.
- Le gouvernement n'a pas de politiques ou de pratiques établies en matière de gestion ou de gouvernance des données, ni d'expérience dans la réalisation d'évaluations des risques liés aux données et à l'IA, ce qui serait nécessaire pour développer l'initiative de manière responsable.

Ces domaines de défis et d'opportunités ont mené à un certain nombre de conclusions pour les prochaines étapes du parcours de partage de données du Jigawa. Un accent sur un projet pilote de cas d'usage petit et gérable, qui peut être mis en œuvre compte tenu des capacités et de la disponibilité des données existantes, est approprié. Cet accent sur la réalisation d'un petit projet pilote pourrait également être utilisé comme un exemple pour sensibiliser et comme un point de départ pour des politiques à l'échelle de l'État sur la gouvernance et la gestion des données.

Cas d'usage

Nous avons organisé un atelier de priorisation des secteurs où nous avons parcouru la feuille de route pour identifier le secteur le plus approprié pour un cas d'usage unique. Ici, le Jigawa avait un certain nombre de secteurs — l'agriculture, la santé et l'éducation — qui semblaient réalisables : ils étaient stratégiquement importants et bénéficiaient d'un fort engagement des parties prenantes. À la suite d'un atelier discutant des facteurs les plus importants pour le choix du secteur, le Jigawa a mené une évaluation de chaque secteur à l'aide de la liste de contrôle.

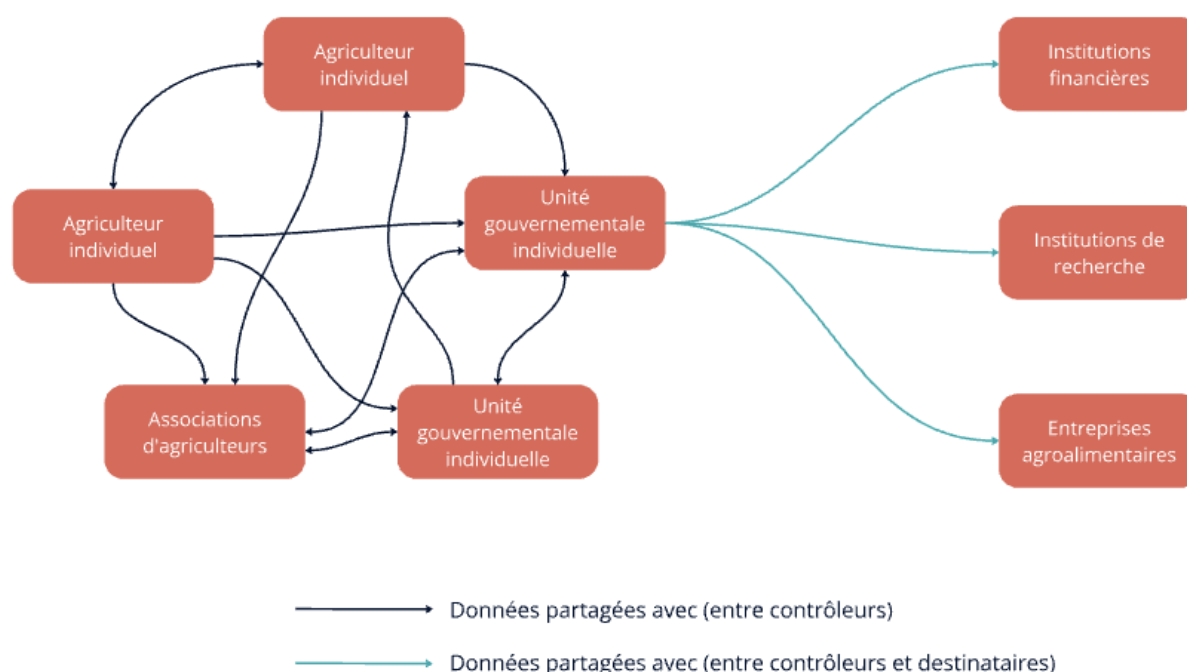
Cependant, chaque secteur a montré un potentiel et a bénéficié d'un fort soutien des équipes sectorielles respectives.

Par conséquent, nous avons co-produit un document passant en revue les caractéristiques spécifiques des cas d'usage potentiels au sein de chaque secteur, compte tenu des projets numériques en cours et des données disponibles dans chaque secteur. Nous avons identifié les problèmes ciblés spécifiques, les données pertinentes à partager, les applications d'IA possibles et les destinataires des données. Sur la base de cette spécification, il est devenu clair que le cas d'usage agricole était le plus avancé en termes de collecte de données, de développement d'infrastructures et de mise en place de canaux de communication entre les parties prenantes. Compte tenu des conclusions mentionnées précédemment lors de l'étape de l'évaluation initiale du pilote, l'équipe a choisi de se concentrer sur le cas d'usage agricole pour le reste du projet pilote.

Dans l'atelier sur le cas d'usage, des domaines d'action future ont été identifiés avant de mettre en œuvre le cas d'usage. Ceux-ci comprenaient l'engagement avec des collègues des équipes juridiques et de l'agence nigériane de protection des données pour recevoir des conseils sur la manière dont la législation s'applique à leurs cas d'usage, l'examen des risques impliqués dans le cas d'usage agricole, et l'engagement avec les acteurs du secteur agricole pour comprendre les données spécifiques qui les intéressent et quelles seraient leurs exigences d'accès.

Mécanisme et gouvernance

Nous avons organisé un atelier avec le Jigawa pour identifier les mécanismes de partage de données pertinents. Nous avons effectué une cartographie plus détaillée des flux de données et exploré les questions liées à la confiance, à la sensibilité des données, au rôle du gouvernement et au niveau d'implication des contrôleurs de données.





Cela a conduit l'équipe à opter pour un modèle hybride de gestionnaire de données (data steward) et de coopérative de données, qui convenait à leurs méthodes de travail existantes avec les parties prenantes externes ainsi qu'avec d'autres unités gouvernementales. Le mécanisme impliquait qu'une seule unité gouvernementale coordonne l'initiative, mais en impliquant un certain nombre d'unités gouvernementales différentes dans la prise de décision concernant les données partagées, avec qui et dans quel but, ainsi que l'utilisation de forums existants pour dialoguer régulièrement avec les associations d'agriculteurs.

Enfin, nous avons organisé une session individuelle sur la gouvernance avec le Jigawa où nous avons discuté de ce à quoi la gestion de ce mécanisme pourrait ressembler en pratique. Les conclusions de cette dernière étape étaient que les prochaines étapes comprenaient l'engagement avec les utilisateurs de données, les contrôleurs de données, les régulateurs et l'équipe de développement d'API afin de rechercher un accord sur les rôles au sein de l'initiative.

Étude de cas : Türkiye

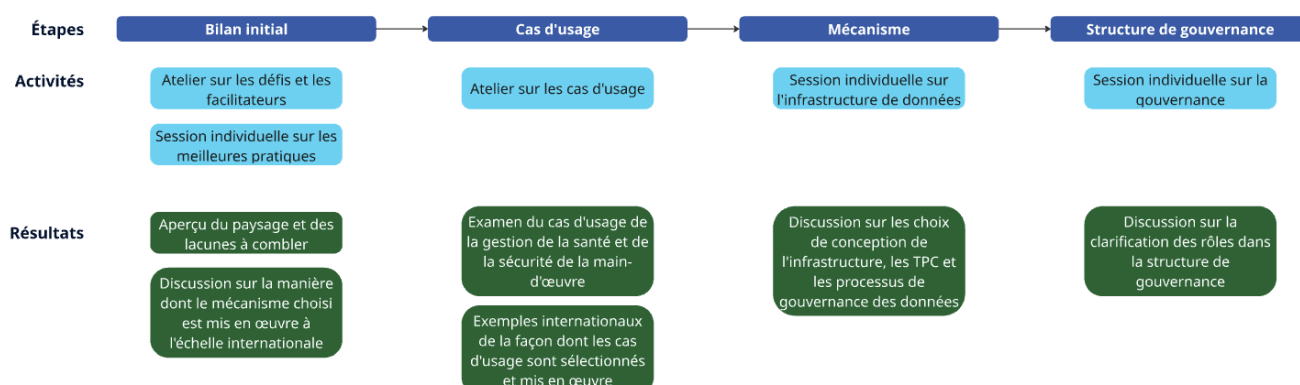
Motivations et point de départ

L'ambition de la Türkiye était de faire progresser ses travaux sur le développement d'un [Espace de données du secteur public](#) (PSDS) et la mise en œuvre de sa [Stratégie Nationale de Données](#) et de sa [Stratégie nationale d'IA](#). Leur PSDS visait à améliorer l'accès à des données gouvernementales de haute qualité, à développer une infrastructure de partage de données plus robuste au sein du gouvernement et à améliorer la culture de gouvernance des données parmi les institutions publiques.

Le Bureau de la transformation numérique de la Présidence de la République de Türkiye (DTO) superviserait les progrès. Ils ont un rôle global dans la gouvernance des données au sein du gouvernement en coordonnant et en guidant les autres institutions et organisations pour qu'elles s'alignent sur des objectifs stratégiques communs. L'Institut statistique turc (TurkStat) est le principal responsable de la mise en œuvre technique du PSDS, agissant en tant qu'opérateur du PSDS. Les travaux sur ce projet avaient déjà commencé et ce projet pilote visait à résoudre un certain nombre de défis auxquels ils étaient confrontés et à fournir un aperçu des meilleures pratiques pour la structure et la gouvernance des espaces de données.

Activités pilotes et résultats

Dans ce projet pilote, le DTO était le principal point de contact. Aux côtés de TurkStat, de l'Autorité de protection des données personnelles et du ministère de l'Industrie et de la Technologie, ils formaient l'équipe de base responsable de l'examen du travail (y compris des listes de contrôle) préparé par Oxford Insights. Certains ateliers ont également été suivis par des parties prenantes plus larges de plusieurs ministères gouvernementaux, notamment TurkStat, l'Autorité de protection des données personnelles, le ministère de la Santé, le ministère de l'Environnement, de l'Urbanisation et du Changement climatique, l'Institution de sécurité sociale (SGK), le Centre d'information et de surveillance des assurances (SGM) et le ministère du Travail et de la Sécurité sociale. Nous avons organisé deux ateliers avec les parties prenantes du DTO ainsi que trois sessions individuelles détaillées avec des experts d'Oxford Insights.



Évaluation Initiale

Le projet pilote a débuté par un atelier sur les défis et les facilitateurs, auquel ont participé plusieurs ministères du gouvernement, qui ont chacun présenté les pratiques de partage et de gouvernance des données qu'ils ont adoptées dans leurs départements. Comme le projet était déjà en cours, cet exercice a été utile pour sensibiliser les parties prenantes intergouvernementales au projet. Il a produit une vue d'ensemble du paysage du partage de données à travers le gouvernement, des différents niveaux de progrès entre les départements, et a aidé à identifier les principales lacunes à combler. Quelques domaines de défis et d'opportunités ont émergé.

Premièrement, jusqu'au projet PSDS, le partage de données avait été largement coordonné au niveau des ministères, ce qui signifiait que différentes normes de données avaient émergé à travers le gouvernement. Le projet PSDS devait aider les fournisseurs de données au sein du projet à rendre les données utilisables dans l'ensemble du gouvernement. Ce travail était déjà soutenu par le projet du Dictionnaire National des Données géré par TurkStat, qui visait à produire des catalogues de données et à standardiser les actifs de données.

Deuxièmement, la manière dont le projet PSDS a été mené présentait un avantage : l'équipe multidisciplinaire comprenait des experts juridiques qui soutenaient une approche de "confidentialité dès la conception" pour les décisions d'architecture des données dès le début. En particulier, cela avait conduit à ce qu'ils optent pour le partage de micro-ensembles de données au sein du PSDS. Cependant, un défi qui subsistait était d'explorer comment l'équipe pourrait utiliser d'autres nouvelles techniques de préservation de la confidentialité, telles que les techniques d'apprentissage fédéré, qui sont promues dans la loi sur l'IA de l'Union européenne.

En nous appuyant sur cet atelier, nous avons tenu une session individuelle sur les meilleures pratiques. Ensemble, nous avons examiné spécifiquement certains des défis identifiés par l'équipe du PSDS, et avons utilisé la norme d'espace de données de l'UE comme exemple de la manière dont les architectures techniques visent à résoudre les problèmes de confidentialité, d'interopérabilité et de confiance au sein des espaces de données. L'introduction d'exemples internationaux de projets connexes a été utile pour l'équipe du DTO afin de voir comment d'autres réagissaient à des défis similaires.

Cas d'usage

L'équipe du DTO avait déjà choisi un cas d'usage pilote pour son PSDS : le partage des données de gestion de la santé et de la sécurité de la main-d'œuvre sur le PSDS. Nous avons organisé un atelier avec le ministère du Travail et de la Sécurité sociale, où ils ont présenté comment ce cas d'usage faciliterait de meilleures prédictions des niveaux de risque d'accident du travail à l'aide de l'intelligence artificielle. À l'aide de la liste de contrôle de la feuille de route, l'équipe du DTO a identifié les risques potentiels qui pourraient être rencontrés en élargissant le scénario du cas d'usage à différentes institutions publiques ainsi que la nécessité de clarifications supplémentaires au sein de l'équipe concernant le partage sécurisé des données et les options de mécanismes d'anonymisation.

Étant donné que le PSDS fonctionnera comme une infrastructure de partage de données réutilisable pour de nombreux cas d'usage, nous avons également examiné à cette étape comment d'autres projets similaires sélectionnent et priorisent les cas d'usage à l'échelle internationale. En particulier, dans le paysage du partage de données du Royaume-Uni, nous avons examiné le Service de données intégrées (IDS) et le programme Better Outcomes through Linked Data (BOLD). Cette session a soulevé des questions clés relatives à la propriété des données, à la manière dont elle est définie et à la façon dont elle évolue tout au long de la chaîne de publication des données au sein du PSDS. Ces questions ont été examinées lors de l'étape subséquente du partenariat pilote.

Mécanisme et gouvernance

Puisque le DTO avait déjà établi les grandes lignes de son mécanisme de partage de données, cette étape s'est concentrée sur les questions clés d'infrastructure de données et de gouvernance que l'équipe avait, et que la feuille de route pouvait être utilisée pour aborder.

Premièrement, la feuille de route a aidé à clarifier la catégorie de mécanisme à laquelle appartient le PSDS, en identifiant le PSDS en développement comme une sorte de mécanisme de plusieurs-à-un, un-à-plusieurs. Cela nous a donné un cadre pour discuter des rôles spécifiques des différentes parties prenantes travaillant sur le projet PSDS. En particulier, nous avons discuté des responsabilités du DTO, de Turkstat et des fournisseurs de données. De manière connexe, nous avons fourni des considérations sur la façon dont le contrôle des données évolue tout au long du PSDS, et comment cela détermine les accords juridiques nécessaires pour établir la confiance entre les acteurs.

Deuxièmement, la feuille de route, et les directives supplémentaires, ont fourni un contexte sur les technologies de préservation de la confidentialité (TPC) qui pourraient être les plus appropriées et importantes ici, telles que l'utilisation de la confidentialité différentielle, de l'anonymisation, de la pseudonymisation et des données synthétiques. Parallèlement, nous avons discuté des outils disponibles qui pourraient être utilisés pour soutenir l'intégration des données provenant de différentes sources à travers le gouvernement, et la gestion des métadonnées.

Ces discussions ont incité l'équipe du DTO à identifier les défis spécifiques auxquels elle pourrait être confrontée à l'avenir, mais ont également apporté de la clarté sur la manière d'aborder et d'atténuer les risques, les aidant ainsi dans la planification stratégique et les capacités de prise de décision dans le contexte du PSDS. L'équipe du DTO a suggéré que la feuille de route est un outil qui peut être utilisé non pas "une seule fois", mais tout au long des autres phases du projet, en particulier lorsque de nouveaux participants se joignent ou lorsque la fonctionnalité ou la conception du projet change.



Annexe 1 : Ressources supplémentaires

Les ressources complémentaires se rapportent à des sections ou questions spécifiques de la liste de contrôle pour le partage de données pour l'IA. Elles sont mises en évidence tout au long du guide de la liste de contrôle. Les ressources que l'on retrouve dans cette section sont :

- [Explicatif sur les droits sur les données](#)
- [Modèle : Cadre de notation quantitative et matrice de priorisation pour les cas d'usage](#)
- [Évaluation des données prêtes pour l'IA](#)
- [Taxonomie des Mécanismes de Partage de Données](#)
- [Infrastructure de données : Composantes](#)
- [Conception de l'infrastructure de données : Exemples](#)
- [Technologies de préservation de la confidentialité : Introduction](#)



Explicatif sur les droits sur les données

Il est important de déterminer dans quelle mesure les droits sur les données sont établis par la législation sur la protection des données de la juridiction de l'équipe et comment la législation peut être complétée par d'autres politiques pour garantir une protection complète de ces droits.

L'explicatif sert d'introduction au concept des droits sur les données et aux moyens par lesquels les cadres législatifs et politiques peuvent les soutenir.

Explicatif sur les droits sur les données

Que sont les droits sur les données?

Émergeant de la législation sur la protection des données et de la jurisprudence à l'échelle mondiale, le concept des droits sur les données pour les individus et les collectifs est devenu central pour une gouvernance solide des données et pour atteindre la justice en matière de données (GPAI, 2022c).

Les droits sur les données reconnaissent que les données sont souvent "cogénérées". C'est-à-dire qu'elles sont créées par de multiples parties : par exemple, des personnes ou des groupes qui sont le sujet des données, une entreprise qui possède un système d'IA ayant joué un rôle dans la génération des données, ou une entreprise qui a traité les données d'une nouvelle manière.

La cogénération de données donne lieu à certains droits pour les parties contributrices en ce qui concerne l'utilisation des données ou la valeur qu'elles génèrent (ALI-ELI, 2021). L'American Law Institute-European Law Institute (ALI-ELI) a recommandé la reconnaissance de l'ensemble non exhaustif de droits sur les données suivant :

- Accès : droit d'accéder aux données, allant de la lecture des données, au traitement des données, jusqu'à leur portabilité complète.
- Cessation : droit d'exiger qu'une partie cesse certaines activités relatives aux données, y compris le contrôle, le traitement et le transfert des données.
- Rectification : droit d'exiger la correction de données incorrectes ou incomplètes.
- Part économique : un droit exceptionnel à une part économique des profits dérivés de l'utilisation des données.

L'applicabilité de ces droits dépend de la circonstance. Les facteurs qui déterminent l'applicabilité incluent la nature et la portée de la contribution d'une partie, les intérêts légitimes de toutes les parties, les intérêts publics, l'équilibre du pouvoir de négociation entre les parties et le type de droit sur les données (Daten Ethik Kommission, 2019).

Comment évaluer la protection des droits sur les données?

Lorsqu'on examine si les droits sur les données sont protégés dans une juridiction, il est important de considérer comment ils sont inclus dans la législation sur la protection des données. La législation peut inclure des droits explicitement, comme le droit à la portabilité des données, mais les droits peuvent également être activés par des exigences concernant

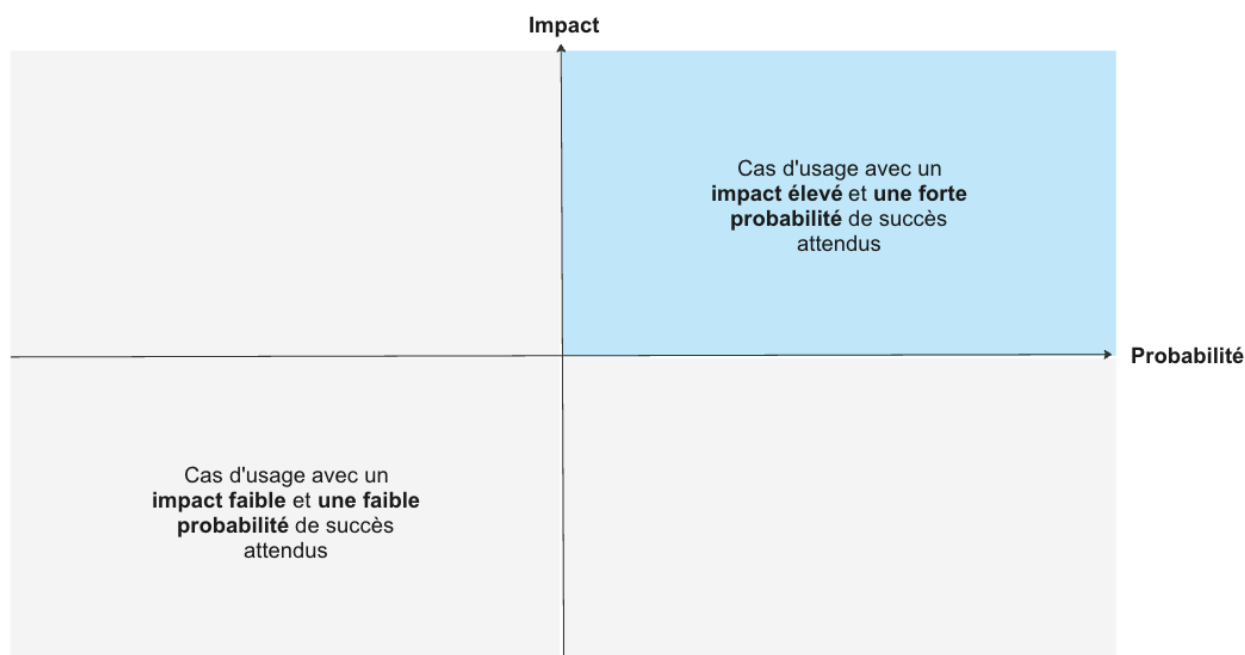


le fait d'être informé de qui détient les données et des transferts de données, de consentir au traitement par des tiers et de se désinscrire du traitement (GPAI, 2023).

Les instruments non juridiques tels que les cadres politiques sont également importants pour soutenir les droits sur les données. Par exemple, le Cadre de politique de données de l'Union africaine contribue au concept du droit à une part économique en appelant les États membres à s'assurer que les Africains bénéficient socialement et économiquement de l'utilisation de leurs données, et que les injustices historiques et les inégalités structurelles ne sont pas perpétuées (GPAI, 2022c). Il est important pour les équipes d'identifier les cadres politiques actuels et futurs au sein de leur gouvernement.

Modèle : Cadre de notation quantitative et matrice de priorisation pour les cas d'usage

Pour aider à comparer les cas d'usage, la liste de contrôle peut être complétée pour différents cas d'usage et un score peut être associé aux résultats. Un modèle de cadre de notation simple est fourni ci-dessous. Les réponses peuvent être notées en fonction du niveau de confiance de l'équipe quant à la conformité du cas d'usage à l'exigence de la liste de contrôle. Le score total est calculé en additionnant les valeurs attachées à chaque réponse. Un score donne une indication globale du potentiel de chaque cas d'usage. Cependant, il y aura probablement des compromis entre les cas d'usage qui sont sélectionnés pour aller de l'avant. Lors de la prise de décision, l'ampleur des bénéfices publics et la probabilité de réussite du projet sont des facteurs importants à prendre en compte, et sont visualisées dans la matrice ci-dessous.



Modèle : Cadre de notation quantitative des cas d'usage



Matrice de décision Impact-Probabilité

		Score de confiance			
Facteurs décisionnels	Question	0 (Non)	1 (Un peu)	2 (Oui, fortement)	3 (Oui, très fortement)
Stratégie & Durabilité	L'initiative cherche-t-elle à résoudre des problèmes clairs qui ont une valeur publique ?				
Communauté & Parties prenantes	Existe-t-il une demande pour les données parmi les utilisateurs de données visés ?				
	L'initiative bénéficie-t-elle du soutien de la ou des équipes pertinentes responsables du secteur ou du domaine dans lequel les données sont partagées?				
	Le public soutient-il le partage de données dans ce secteur?				
Aspects juridiques & Gouvernance	Existe-t-il un cadre juridique qui énonce clairement les conditions de partage des données applicables à cette initiative et respectées par celle-ci?				
	Les données à partager restent-elles dans les limites de ce qui est nécessaire pour le cas d'usage?				
Technologie	Les données partagées dans le cadre de l'initiative peuvent-elles être préparées pour être utilisées par les développeurs d'IA?				
Éthique	Les applications d'IA prévues sont-elles conformes aux directives gouvernementales ou à d'autres pratiques exemplaires en matière d'utilisation et de développement responsables de l'IA?				
	Les risques pour les personnes concernées et les autres groupes affectés peuvent-ils être atténués?				
				Total	

Évaluation des données prêtes pour l'IA

Le tableau ci-dessous présente les quatre facteurs à prendre en compte lors de l'évaluation de la pertinence et de l'utilité d'un ensemble de données pour un cas d'usage (Dama UK, 2013 ; Jolliffe et al., 2023 ; Ravi et al., 2022 ; ESIP Data Readiness Cluster, 2022 ; World Bank, 2019). Comme décrit ci-dessus, cette évaluation peut être adaptée en fonction des données et du cas d'usage. Le tableau fournit des questions visant à évaluer les données dans un cadre de collaboration ou d'atelier avec les participants potentiels de l'initiative.

Questions pour l'évaluation des données			
Qualité	Pertinence pour le cas d'usage	Conformité éthique et légale	Besoins de (pré)traitement
Précision : les données mesurent-elles avec précision le concept du monde réel qu'elles sont censées mesurer? Leurs erreurs sont-elles connues et peuvent-elles être modélisées?	Couverture : les données couvrent-elles l'ensemble de la population d'intérêt pour le cas d'usage pertinent? Pour « couvrir » une population d'intérêt, il suffit qu'elles soient représentatives plutôt que de couvrir une population dans son intégralité.	Confidentialité : la confidentialité des personnes concernées peut-elle être protégée par l'anonymisation et d'autres techniques de préservation de la confidentialité?	Format : les données sont-elles stockées dans un format approprié? Sont-elles lisibles par une machine et faciles à télécharger dans un système IA/ML?
Granularité : les données peuvent-elles être ventilées par sous-groupes pertinents au sein de la population qu'elles mesurent?	Fréquence : les données sont-elles produites et/ou mises à jour à un rythme qui correspond à ce qui est nécessaire pour le cas d'usage?	Conformité réglementaire : les données respectent-elles les lois et réglementations pertinentes régissant le partage et l'utilisation des données dans ce secteur/relatives à ce cas d'usage et à ses données associées? Si la base légale pour le partage de données est le consentement, un consentement approprié a-t-il été recueilli auprès des personnes concernées qui fournissent les données pertinentes?	Annotation : Les données disposent-elles de métadonnées détaillées, lisibles à la fois par l'homme et par la machine?
Intégrité : les sources de données, le	Temporalité : Les données sont-elles	Minimisation et proportionnalité : les	Interopérabilité : les données



Questions pour l'évaluation des données

Qualité	Pertinence pour le cas d'usage	Conformité éthique et légale	Besoins de (pré)traitement
traitement et la gestion de l'accès sont-ils suivis et dignes de confiance?	mises à jour suffisamment rapidement après qu'un événement se produise pour le cas d'usage concerné?	données partagées sont-elles le minimum nécessaire pour remplir les objectifs prévus dans le cas d'usage?	peuvent-elles être liées à d'autres ensembles de données via des identifiants communs et cohérents?
Quantité : y a-t-il un volume de données suffisant pour l'utilisation prévue dans les systèmes d'IA dans le cas d'usage?	Comparabilité : les données sont-elles comparables dans le temps et d'un endroit à l'autre ? Dans quelle mesure les comparaisons peuvent-elles être faites avec des données passées/des données similaires d'ailleurs?	Biais : les données sont-elles exemptes de biais systématiques qui pourraient entraîner des résultats disproportionnellement négatifs pour certains groupes sociaux?	

Taxonomie des mécanismes de partage de données

La taxonomie des mécanismes de partage de données fournit des exemples des trois catégories de mécanismes de partage de données introduites dans le cadre de la feuille de route. La taxonomie divise chaque catégorie en exemples, fournissant une description de l'arrangement de partage de données, des exemples concrets et des détails sur le moment où chacun est couramment utilisé.

- **Partage de données individuel** : accords entre deux parties individuelles, qui sont des utilisateurs et/ou des contrôleurs des données.
- **Partage de données un-à-plusieurs** : accords entre une partie, qui est le contrôleur de données et de nombreux utilisateurs de données. La taxonomie comprend également le partage de données de plusieurs-à-un, un-à-plusieurs. Il s'agit d'une sorte d'arrangement de partage de données un-à-plusieurs, où il y a des accords entre une partie, qui représente les contrôleurs de données, et de nombreux utilisateurs de données.
- **Partage de données plusieurs-à-plusieurs** : accords entre de nombreuses parties qui sont des contrôleurs de données et de nombreuses parties qui sont des utilisateurs de données. Il peut s'agir des mêmes parties.

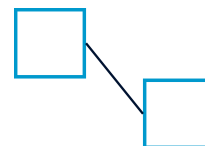
Les diagrammes illustratifs de la taxonomie utilisent la clé ci-dessous. En pratique, le type d'acteurs impliqués dans l'initiative et leurs rôles diffèrent.

Clé de la taxonomie

Exemples de participant	Rôles des participants
 Organisme principal du secteur public	 Contrôleurs de données
 Organisations du secteur privé	 Utilisateurs de données en bleu
 Milieu universitaire et recherche	 Utilisateurs et contrôleurs de données en jaune
 Personnes concernées par les données individuelles	 Gestionnaire de données
	Accord juridique 

Partage de données individuel

Les mécanismes de partage de données individuel établissent une connexion entre le gouvernement et une autre entité par laquelle les données peuvent être partagées. Dans les cas pertinents pour ce projet, le gouvernement est toujours un fournisseur de données ; il peut aussi être un utilisateur de données. Dans ces cas, il y a un partage de données multidirectionnel entre une agence gouvernementale et une autre organisation.



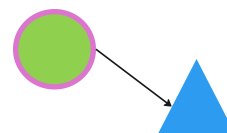
Quand sont-ils les plus pertinents?

Les mécanismes individuels sont utiles lorsque le gouvernement s'approvisionne en recherche ou en exécution d'un projet avec une portée clairement définie. Il permet au gouvernement de s'engager étroitement avec une entité particulière pour résoudre un petit problème bien défini ou mener une recherche, entre autres choses.

Quels sont quelques exemples?

Quels sont quelques exemples?

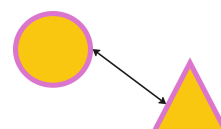
Un accord est conclu entre une agence gouvernementale et une autre organisation. Cela pourrait être pour un projet qui implique le partage de données ou simplement pour les données elles-mêmes..



Exemple concret : Le [Registre de réponse rapide](#) (RRR) est un cadre de réponse aux chocs conçu par le Programme national de filets de sécurité sociale du Nigeria (NASSP) en collaboration avec la Banque mondiale pour fournir des interventions sociales d'urgence aux ménages pauvres et vulnérables dans des circonstances difficiles.

Partenariats public-privé (PPP)

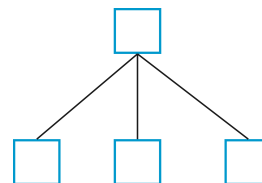
Les PPP impliquent une collaboration entre des agences gouvernementales et des entités privées pour mener ce qui sont généralement des entreprises à grande échelle. Ils sont généralement financés par des fonds privés, la charge des risques reposant sur l'entité privée. Dans des cas comme ceux-ci, les données sont souvent partagées dans les deux sens, entre les deux parties : le gouvernement est à la fois un fournisseur et un utilisateur de données.



Exemple concret : Le partenariat [Vers des transports routiers zéro émission](#) (2zero) est un accord entre la Commission européenne et des partenaires de l'industrie automobile pour accélérer la transition vers un système de transport routier neutre en carbone d'ici 2050.

Un-à-plusieurs

Les mécanismes de partage de données un-à-plusieurs fonctionnent par le biais de multiples accords juridiques entre un seul fournisseur de données et des utilisateurs de données. Dans les cas pertinents, le fournisseur de données sera généralement une agence ou un ministère gouvernemental. Chaque accord avec l'utilisateur de données peut être sur mesure, bien qu'ils puissent aussi être identiques. Dans le cas où ils ne sont pas identiques, ils fonctionnent comme de multiples accords/mécanismes de partage de données un-à-un qui peuvent être regroupés par un mécanisme d'accès partagé pour plus de facilité, malgré les différents accords avec les partenaires.



Quand sont-ils les plus pertinents?

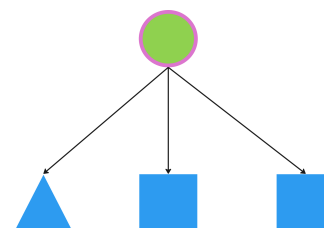
Ils sont utiles lorsque le gouvernement détient des données qu'il souhaite partager avec plusieurs parties différentes. Il peut s'agir d'organisations universitaires, privées ou du tiers secteur, qui peuvent toutes exiger des accords et des conditions d'utilisation différents.

Quels sont quelques exemples?

Données ouvertes

Les données sont accessibles, exploitables et partageables par/à quiconque sous une licence de données ouvertes. Dans ce cas, les accords sont tous les mêmes malgré leur utilisation par de multiples parties différentes.

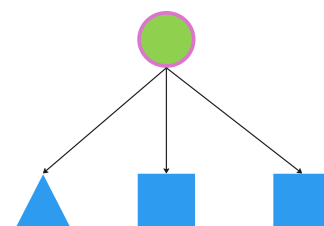
Exemple concret : datos.gob.cl, la plateforme de données ouvertes du Chili, est un répertoire de données où les organismes gouvernementaux publient leurs données sous une licence de données ouvertes.



Bac à sable (Sandbox)

Un bac à sable de données est un environnement sécurisé qui permet aux utilisateurs externes au gouvernement d'accéder et d'utiliser l'ensemble de données d'une manière qui n'affecte pas l'intégrité de l'ensemble de données original ni ne compromet la sécurité des données.

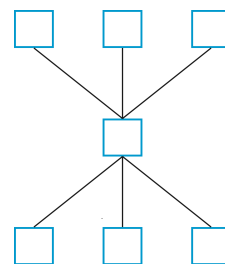
Exemple concret : Le [Bac à sable Réglementaire de Données Singapourien](#) fournit aux développeurs de produits un environnement contrôlé dans lequel développer leurs technologies axées sur les données, en les testant par rapport aux lois existantes sur la protection des données.



Plusieurs-à-un, un-à-plusieurs

Les cas de mécanismes de partage de données de plusieurs-à-un, un-à-plusieurs semblent imiter les relations plusieurs-à-plusieurs, mais sont subtilement différents et nécessitent des structures de gouvernance et des cadres juridiques différents.

Ils contiennent un nœud central, représenté ici comme une entité gouvernementale, bien que ce ne soit pas toujours le cas, qui collecte des données auprès d'un certain nombre de fournisseurs de données différents et les distribue ensuite à de multiples utilisateurs de données différents. Les fournisseurs et les utilisateurs de données agissent par l'intermédiaire de ce nœud et n'ont des droits et des responsabilités juridiques que vis-à-vis de cette entité. Dans les mécanismes de partage de données de plusieurs-à-un, un-à-plusieurs, les fournisseurs et les utilisateurs ne s'engagent pas dans un accord direct.



Quand sont-ils les plus pertinents?

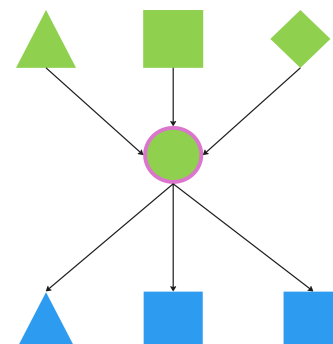
Ils sont utiles lorsqu'il existe des réseaux complexes d'utilisateurs et de fournisseurs de données et qu'il leur est difficile de s'engager directement. Les mécanismes de plusieurs-à-un, un-à-plusieurs permettent à un intermédiaire d'agir comme un conduit, rationalisant les interactions. Au lieu que tous les fournisseurs de données tentent d'interagir avec tous les utilisateurs de données, ils interagissent chacun avec un intermédiaire désigné. Comme mentionné, cet intermédiaire n'a pas besoin d'être une entité gouvernementale. Il pourrait tout aussi bien être un tiers ou un véhicule spécial qui est mis en place pour être le nœud intermédiaire du diagramme ci-dessus.

Quels sont quelques exemples?

Fonds de données (Data trust)

Un fonds de données est une entité/structure juridique où les fournisseurs/sujets de données accordent une autorité à une organisation désignée, parfois une organisation tierce mise en place spécifiquement pour cela, pour prendre des décisions concernant le partage de leurs données avec d'autres entités.

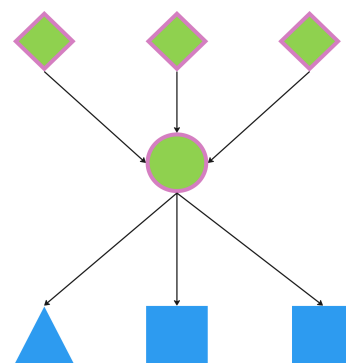
Exemple concret : Le [Fonds de Données pour l'Action Climatique](#) est un référentiel basé sur la blockchain de toutes les principales données de registre de crédits carbone pour améliorer la transparence comptable conformément à l'article 6 de l'Accord de Paris.



Gestionnaire de données (Data steward)

La gestion de données est un concept participatif d'intermédiaires de données où des acteurs tiers, des gouvernements ou des ONG, par exemple, reçoivent l'autorité des personnes concernées par les données pour faciliter l'accès direct aux données qu'elles produisent par des parties qui veulent utiliser ces données.

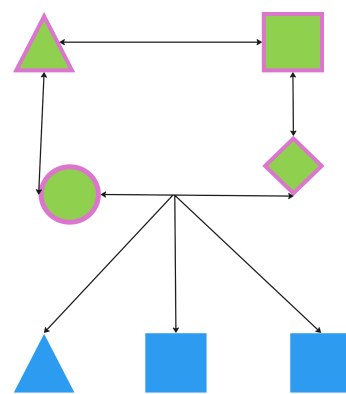
Exemple concret : Le Carnet de santé à Taiwan où le gouvernement aide à faciliter l'accès par des tiers aux données de santé des patients.



Coopérative de données

Les membres d'une communauté contribuent conjointement à leurs données mises en commun et les gèrent. Tous les membres conservent leurs droits sur les données.

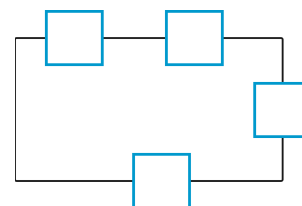
Exemple concret : [MIDATA](#), une coopérative de données suisse créée en 2015, donne aux utilisateurs un contrôle total sur leurs données de santé. D'autres exemples de coopératives de données sont la [Scottish Agricultural Organisation Society](#) (SAOS) ou [POSUMO](#), une autre coopérative suisse pour les données de mobilité.





Plusieurs-à-plusieurs

Les cas de mécanismes de partage de données plusieurs-à-plusieurs semblent imiter ceux des mécanismes plusieurs-à-un, un-à-plusieurs illustrés ci-dessus. Ils sont cependant subtilement différents et nécessitent donc des structures de gouvernance différentes.



Les mécanismes de partage de données plusieurs-à-plusieurs établissent des processus de partage de données multipartites et multidirectionnels, où les données circulent souvent avec des processus clairs pour l'accès au sein de certains acteurs prédéterminés qui peuvent être à la fois des fournisseurs et des utilisateurs. Ce partage peut être facilité par une plateforme particulière, mais il n'existe pas de nœud central par lequel passe toute la collecte de données et le partage ultérieur ; au lieu de cela, de nombreuses parties s'engagent avec de nombreuses autres parties simultanément.

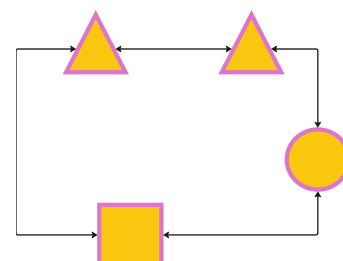
Quand sont-ils les plus pertinents ?

Ils sont utiles lorsqu'il existe des réseaux complexes de parties prenantes où un engagement direct peut être facilement facilité, comme dans le cas d'un marché de données plusieurs-à-plusieurs, ou lorsqu'il y a de nombreuses parties au sein d'un écosystème qui sont déjà au courant les unes des autres. De tels mécanismes permettent aux parties au sein de la structure de partage de données de s'engager les unes avec les autres et d'exercer leur pouvoir sur leurs avoirs/besoins en données particuliers.

Quels sont quelques exemples ?

Marché de données (Data marketplace)

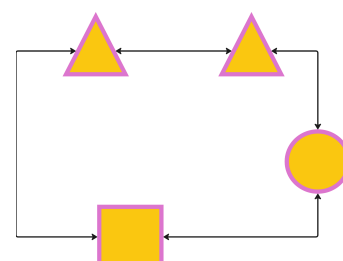
Un marché de données : plusieurs-à-plusieurs fonctionne légèrement différemment de celui mentionné ci-dessus. Au lieu qu'un intermédiaire collecte toutes les données des fournisseurs et les distribue ensuite aux utilisateurs, il agit simplement comme une plateforme pour faciliter l'échange. Dans ce cas, les fournisseurs et les utilisateurs s'engagent directement les uns avec les autres en utilisant la plateforme fournie.



Exemple concret : [Dawex](#), un marché qui permet aux acteurs du monde entier de se procurer, d'échanger et de monétiser des données

Communauté de données (Data commons)

Communauté de données : Les membres d'une communauté contribuent conjointement à leurs données mises en commun, les gèrent, les partagent et les utilisent. Ils peuvent utiliser une plateforme cloud ou un autre environnement sécurisé pour faciliter l'accès restreint aux membres pertinents de la communauté.



Exemple concret : [Aclimate](#), une plateforme de données agricoles en Colombie.

Infrastructure de données : Composantes

Le tableau ci-dessous décompose les composantes de l'infrastructure de données d'une initiative de partage de données et illustre les options de conception d'infrastructure que l'équipe développe.

Infrastructure de données	
Composante de l'infrastructure de données	Exemple options et considérations
Stockage des données : où les données sont stockées.	Il existe trois types d'architecture pour le stockage des données sur lesquels l'infrastructure de partage de données peut s'appuyer (soit un, soit un hybride) : ● Centralisée : toutes les données résident dans un seul référentiel, souvent un entrepôt de données ou une plateforme de données d'entreprise. ● Fédérée : les données restent distribuées sur plusieurs systèmes, chacun ayant sa propre autonomie. ● Décentralisée : les données sont distribuées sur un réseau de nœuds et sans autorité centrale. Chacun a des avantages différents. L'architecture centralisée peut être utilisée lorsqu'il y a, ou peut y avoir, un seul contrôleur des données. L'architecture fédérée peut convenir lorsque l'accès inter-système est nécessaire, mais qu'il y a plusieurs contrôleurs de données et que le maintien de l'autonomie est important. L'architecture décentralisée peut être utilisée lorsque la sécurité, la confidentialité et la transparence sont essentielles au sein d'un écosystème de partage de données de pair à pair Qu'en est-il des architectures hybrides ? Les architectures hybrides peuvent permettre à une organisation de tirer parti des forces de chaque type d'architecture tout en répondant aux exigences spécifiques en matière d'analyse, de science des données et de conformité. Par exemple, une organisation peut utiliser une architecture hybride qui combine un entrepôt de données centralisé pour les données structurées, un lac de données pour les données non structurées et une chaîne de blocs pour les registres de transactions sécurisés. Cela peut être réalisé grâce à un mélange de technologies comme Hadoop, Snowflake et la blockchain pour prendre en charge les différents types de données et cas d'usage. Il y a quelques autres considérations liées au stockage et au traitement des données que les organisations peuvent vouloir prendre en compte : a) Dans le cadre d'une approche décentralisée, les organisations peuvent également mettre en œuvre un « data mesh ». Cela permet de donner aux équipes spécifiques à un domaine les moyens de gérer leurs propres produits de données.



Infrastructure de données

Composante de l'infrastructure de données	Exemple options et considérations
	b) L'architecture Lambda peut être utilisée pour gérer le traitement des données. Cette architecture combine le traitement par lots et en temps réel pour l'analyse et les rapports. L'architecture Kappa peut également être considérée. Elle est similaire à Lambda, mais repose uniquement sur le traitement des flux pour la gestion des données en temps réel. Les données peuvent être stockées sur site, sur le cloud ou dans un système hybride. Cela dépendra des considérations de sécurité et de la conformité à la législation et aux réglementations applicables en matière d'hébergement de données.
Intégration des données : les sources de données et la manière dont elles sont ingérées dans l'installation de stockage des données, y compris les transformations qui doivent avoir lieu et les normes qu'elles suivent.	La manière dont l'intégration des données varie, naturellement, selon l'architecture. L'architecture centralisée dépend des pipelines ETL (extraction, transformation, chargement) ou de l'entreposage de données. Les données provenant de différentes sources telles que les bases de données, les applications, les fichiers, les API et le stockage cloud peuvent être ingérées en mode par lots, par diffusion en continu en temps réel, ou les deux. Les transformations sont effectuées par le biais du nettoyage, de la normalisation, de l'enrichissement et de l'agrégation selon les normes : SQL, XML, JSON, CSV. L'architecture fédérée permet d'accéder à de nombreuses bases de données, entrepôts de données, applications et services cloud à l'aide de moteurs de requête fédérés ou de plateformes de "data fabric". Les données sont accédées sur place par virtualisation ou réplication avec un minimum de changements, bien que ceux-ci se produisent fréquemment de manière dynamique au moment de la requête, en utilisant des normes telles que SQL, les API REST ou GraphQL. L'intégration de données décentralisée utilise des plateformes blockchain et des marchés de données décentralisés ; les sources de données sont des appareils IoT, des capteurs, des médias sociaux et des DApps. L'ingestion se fait par le biais de partage pair à pair avec des contrats intelligents, tandis que l'étendue du travail de transformation effectué à la source est minimale, alors que les normes à suivre sont des formats de données spécifiés par la blockchain et des DID. Indépendamment de l'architecture, la qualité des données doit être assurée par le profilage, le nettoyage et la validation. De plus, les référentiels de métadonnées gérés de manière centralisée aident à comprendre la lignée des données, les usages et la qualité. Des politiques de gouvernance des



Infrastructure de données

Composante de l'infrastructure de données	Exemple options et considérations
	données solides, des processus et des contrôles d'accès aux données, des protocoles de sécurité et de conformité doivent être présents dans toutes les architectures. Par exemple, le processus d'intégration des données de vente charge les données de divers systèmes transactionnels dans un entrepôt de données central ; il extrait d'abord les données, puis les transforme en les nettoyant, en les normalisant et en les agrégeant, et charge finalement ces données transformées des systèmes sources dans un entrepôt de données correctement formaté où elles sont prêtes à être utilisées pour l'analyse et les rapports.
Découverte des données : comment les utilisateurs découvrent quelles données sont disponibles et les informations sur les données.	Outils pour rechercher les données. Ceux-ci incluent les catalogues de données (Collibra, Alation, IBM watson catalogue).
Accès aux données : comment l'accès des utilisateurs est géré et activé.	La méthode d'accès aux données dépend également du choix de stockage des données, car cela déterminera la manière dont les requêtes sont exécutées : ● Centralisé : l'accès aux données est activé via un seul point d'entrée. ● Fédéré : une interface unifiée ou un moteur de requête fédéré fournit un accès aux données sur plusieurs systèmes. ● Décentralisé : partage pair à pair et interaction directe entre les nœuds. L'accès aux données peut être activé via des API qui acheminent les données en toute sécurité vers les utilisateurs depuis le stockage. L'utilisation des API est courante si les données sont consultées en externe ou si un modèle d'apprentissage automatique est en cours d'exécution pour extraire automatiquement les données. Alternativement, des transferts de données par lots peuvent être utilisés si les données sont mises à jour peu fréquemment. Si les utilisateurs utilisent des API au sein d'une architecture fédérée, un « data fabric » peut être nécessaire pour intégrer les données de différentes sources de données. En plus de la manière dont l'accès aux données est activé, des contrôles d'accès aux données sont utilisés pour déterminer quelles requêtes



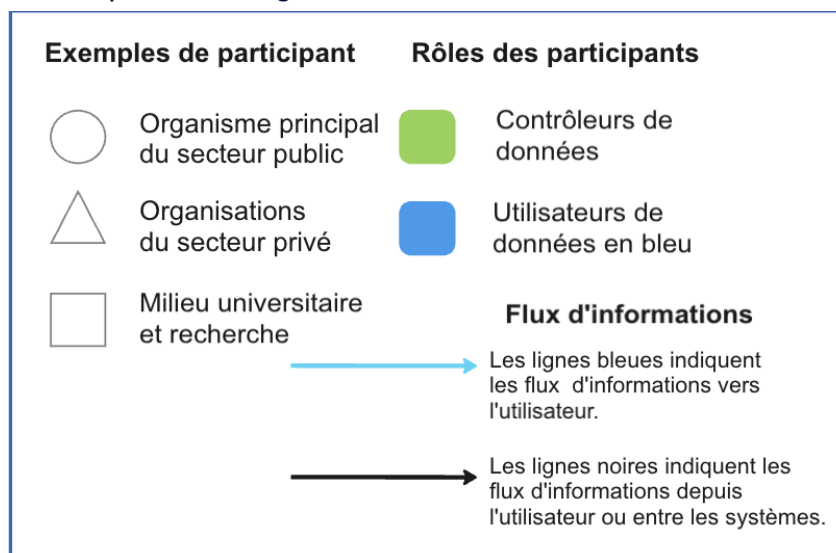
Infrastructure de données

Composante de l'infrastructure de données	Exemple options et considérations
	d'utilisateur sont autorisées. Les types de contrôles comprennent le Contrôle d'accès basé sur les rôles, le Contrôle d'accès discrétionnaire (DAC), le Contrôle d'accès obligatoire (MAC), entre autres. Le choix du contrôle dépend de la catégorie des utilisateurs et de la sensibilité des données. Les données plus sensibles nécessitent des contrôles plus stricts (par exemple, MAC, sécurité axée sur les données).
Services de données : autres services que l'initiative fournit aux utilisateurs de données, tels que des visualisations, des analyses et des applications.	Les services de données fournissent l'interface pour l'extraction, la transformation et le chargement des données au sein d'une conception architecturale qui peut se faire via des interactions basées sur une interface graphique (GUI) et non-graphique (non-GUI). Dans les architectures centralisées, les tableaux de bord de BI offrent des interfaces visuelles pour explorer et analyser les données extraites via des API depuis le référentiel central (par exemple, un entrepôt de données), tandis que les interfaces non-GUI comme les clients SQL permettent des requêtes directes. Dans les architectures fédérées, les tableaux de bord de BI peuvent intégrer des données provenant de plusieurs sources via des moteurs de requête fédérés, tandis que l'accès programmatique est possible via des API ou des interfaces de « data fabric ». Dans les architectures décentralisées, les tableaux de bord de BI pourraient visualiser les données des contrats intelligents, tandis que les interfaces non-GUI comme les explorateurs de blockchain ou les scripts personnalisés interagissent directement avec la blockchain. Tous les services ou applications supplémentaires sont susceptibles de dépendre des besoins des utilisateurs, des objectifs de l'initiative et des contraintes de ressources de l'équipe.

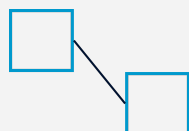
Conception de l'infrastructure de données : Exemples

Alors que l'infrastructure de données variera selon chaque gouvernement et cas d'usage, quelques grandes conceptions d'infrastructure de données à travers différents types d'arrangements de partage de données sont illustrées ci-dessous.

Clé pour les diagrammes d'infrastructure de données



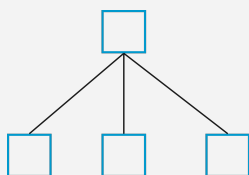
Arrangement de partage de données individuel



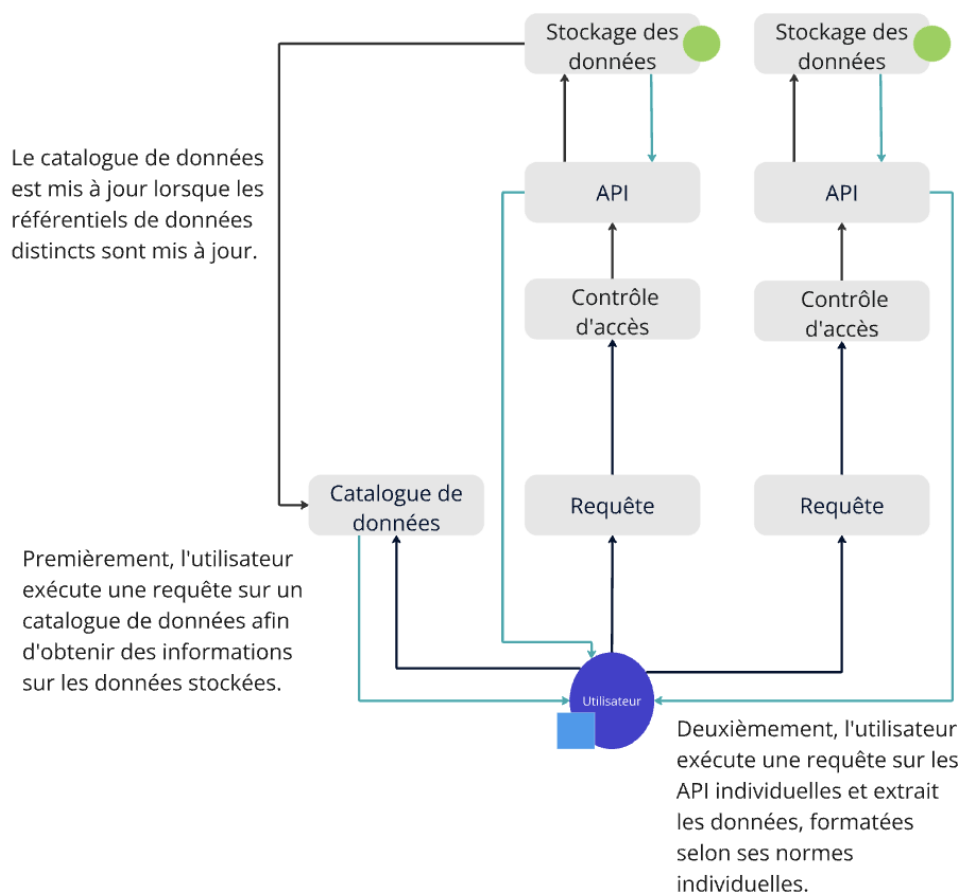
Un référentiel de données unique, par exemple un entrepôt de données, auquel l'utilisateur accède.



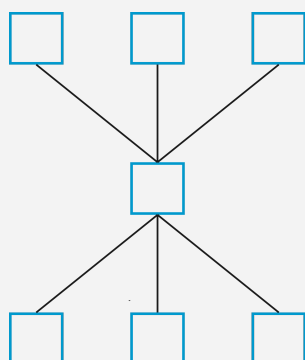
Arrangement de partage de données un-à-plusieurs



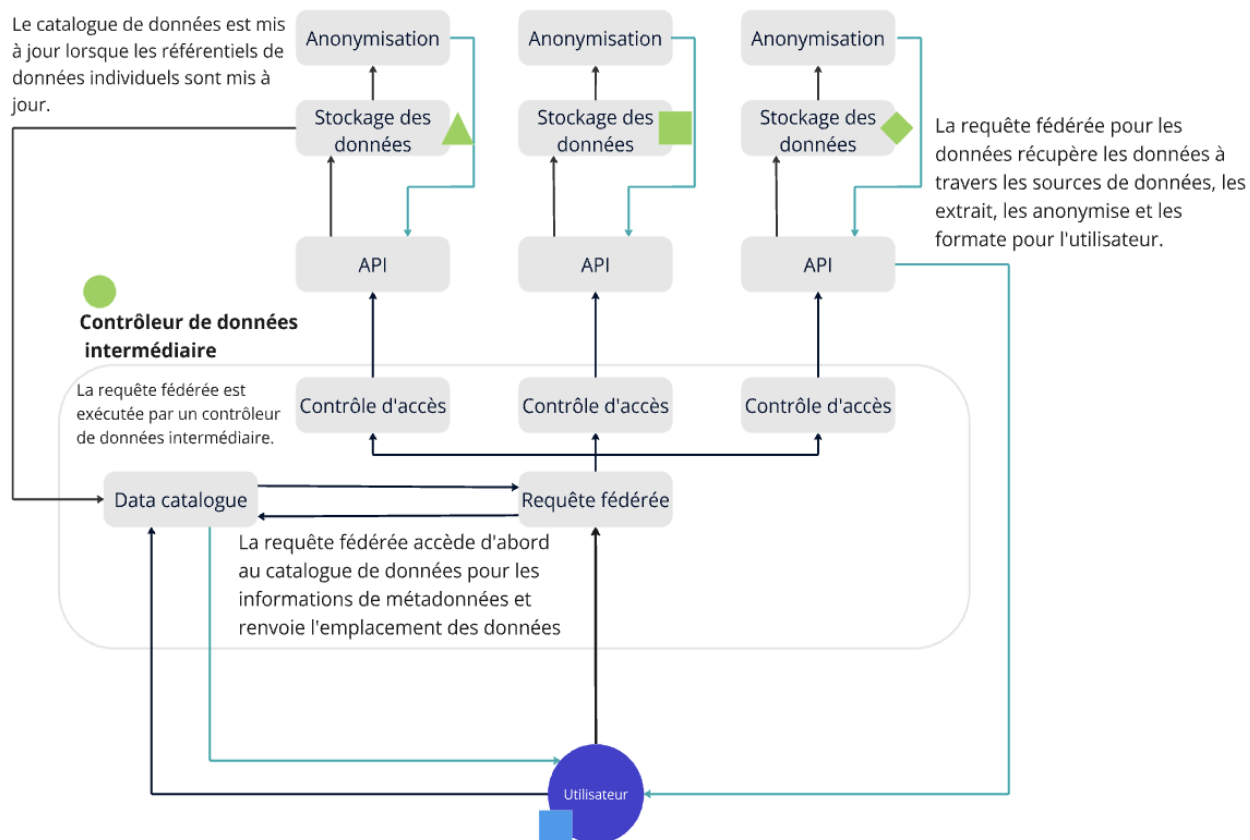
Un contrôleur de données, par exemple une agence gouvernementale, met les données qu'il détient à la disposition des utilisateurs de données. Dans le cas ci-dessous, l'agence détient des données dans différents référentiels de données.



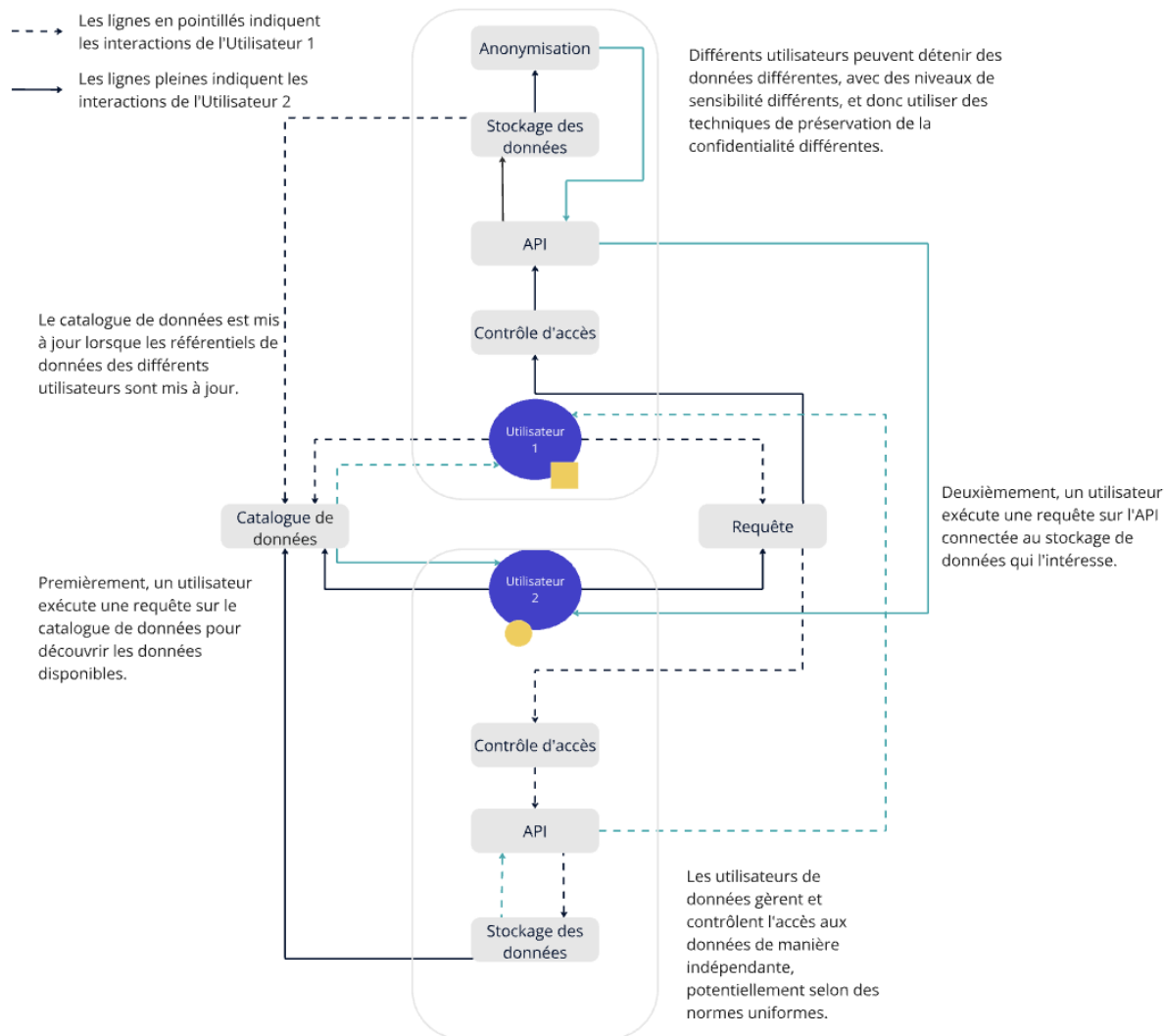
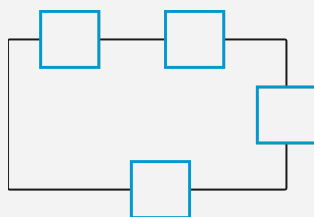
Arrangement de partage de données plusieurs-à-un, un-à-plusieurs



Le catalogue de données est mis à jour lorsque les référentiels de données individuels sont mis à jour.



Arrangement de partage de données plusieurs-à-plusieurs





Technologies de préservation de la confidentialité : Introduction

Avant de s'engager avec les TPC, il y a un certain nombre de procédures générales qui peuvent être envisagées pour maintenir la confidentialité des données lors du partage de données. Premièrement, il est important de se demander si le partage de données est approprié dans le cas d'usage pertinent. Deuxièmement, les équipes peuvent envisager de partager des données à un niveau d'abstraction plus élevé qui maintient la confidentialité des personnes concernées : ne partager que des données au niveau du groupe, par exemple.

Si aucune de ces procédures générales n'est appropriée, par exemple si des données sont nécessaires et que des données au niveau individuel sont requises pour le cas d'usage particulier, les équipes peuvent alors envisager un certain nombre de TPC différentes. Voici un ensemble d'exemples illustratifs de techniques que vous pouvez utiliser.

Technique	Comment ça fonctionne	Quand est-ce pertinent?	Considérations supplémentaires
Anonymisation/ pseudonymisation	L'anonymisation supprime les informations personnellement identifiables (IPI) des ensembles de données, ce qui rend impossible de remonter à un individu. La pseudonymisation remplace ces IPI par des pseudonymes, ce qui permet de remonter aux données d'origine uniquement avec des informations confidentielles supplémentaires.	Ensembles de données où les identifiants directs doivent être protégés et scénarios nécessitant une conformité aux réglementations sur la protection des données comme le RGPD.	L'anonymisation peut être difficile à réaliser parfaitement, car il y a toujours un risque de réidentification, surtout avec de grands ensembles de données ou lorsqu'elle est combinée avec d'autres sources d'information. La pseudonymisation est souvent considérée comme une approche plus pratique pour de nombreux cas d'usage.
Confidentialité différentielle	Ajoute du bruit statistique aux données/requêtes pour protéger les points de données individuels.	C'est pertinent pour les grands ensembles de données où les points de données individuels sont sensibles et où des informations agrégées sont nécessaires sans révéler des détails spécifiques.	Le niveau de bruit ajouté dans la confidentialité différentielle est un compromis entre la confidentialité et la précision des résultats : des niveaux de confidentialité plus élevés peuvent entraîner des informations agrégées moins précises ou utiles.
Chiffrement homomorphe	Le chiffrement homomorphe permet d'effectuer des calculs sur des données chiffrées sans avoir besoin de les déchiffrer. Les résultats, une fois déchiffrés,	Pour les données personnelles sensibles (par exemple, les dossiers de santé ou les données financières) et les situations où l'analyse doit être effectuée sans	Cela peut être gourmand en calcul, surtout pour les calculs complexes. Cela peut être un facteur limitant dans certaines applications.



Technique	Comment ça fonctionne	Quand est-ce pertinent?	Considérations supplémentaires
	correspondent aux résultats des opérations effectuées sur le texte en clair.	révéler les données à des tiers.	
Données synthétiques	Les données synthétiques sont des données générées artificiellement qui imitent les propriétés (statistiques) de données réelles pertinentes sans contenir d'informations personnelles réelles. Ces données sont créées à l'aide de modèles qui sont entraînés sur des données réelles. C'est une option de plus en plus populaire pour les gouvernements, et une étude de cas du Royaume-Uni est explorée ci-dessous.	Tous les scénarios où les données réelles sont trop sensibles pour être partagées et les applications nécessitant de grands ensembles de données pour les tests et le développement sans risquer la confidentialité.	La qualité et l'utilité des données synthétiques dépendent fortement des modèles sous-jacents utilisés pour les générer. Elles peuvent ne pas toujours être un substitut parfait aux données réelles dans tous les scénarios.
Apprentissage fédéré	Les modèles d'apprentissage automatique sont entraînés sur plusieurs appareils/serveurs fédérés détenant des échantillons de données locales, sans les échanger ; seules les mises à jour du modèle sont partagées pour améliorer un modèle central.	Lorsque des ensembles de données distribués ne peuvent pas être centralisés en raison de préoccupations de confidentialité et dans les cas où les données locales pourraient être utilisées pour améliorer un modèle global sans partager de données brutes.	L'apprentissage fédéré peut introduire des défis en termes de convergence du modèle et de garantie de performances constantes sur divers ensembles de données détenus par différentes parties.
Calcul multipartite sécurisé (MPC)	Le MPC permet à plusieurs parties différentes de calculer conjointement une fonction sur leurs entrées tout en gardant ces entrées privées. Chaque partie n'apprend que le résultat de chaque calcul, mais pas les entrées des autres parties.	Cas d'engagement multipartite avec des données sensibles et tâches d'analyse de données conjointes où les données d'entrée de toutes les parties doivent rester confidentielles.	Il peut être gourmand en calcul, surtout pour les calculs complexes impliquant de nombreuses parties.

Exemple : Projet Pilote de Données Synthétiques du NHS en Angleterre

Les données de santé sont extrêmement importantes : leur analyse permet aux chercheurs et aux cliniciens de mieux comprendre les maladies existantes, les traitements et la prestation de services, tout en permettant la recherche sur les futurs traitements. Cela dit, elles sont également extrêmement sensibles. Des mesures de confidentialité robustes sont nécessaires pour garantir leur sécurité. L'accès à ces données est donc actuellement un processus assez long, car seuls les utilisateurs approuvés, avec des projets approuvés, sont autorisés à y accéder. Il peut également être difficile pour les utilisateurs potentiels de savoir à l'avance à quel point les données seront utiles ; par exemple, quels ensembles de données seront disponibles, quels champs seront présents dans les ensembles de données, etc.

Le NHS mène un projet pilote pour aider les utilisateurs potentiels à comprendre la forme générale des données de patients réelles grâce à des ensembles de données synthétiques à partir de valeurs approximatives aux colonnes (NHS, n.d). Les données synthétiques générées imitent la structure, le format et le volume des données, mais ne préservent pas, dans ce cas, les relations entre les champs.

Ainsi, bien que ces ensembles de données ne puissent pas être utilisés actuellement pour l'analyse, ils permettront aux utilisateurs de données potentiels de mettre en place l'infrastructure, de tester le code et d'évaluer les pipelines avant d'accéder aux données réelles. Cela permet aux utilisateurs de déterminer si les données sont appropriées sans avoir à partager des données sensibles ou à subir un long processus d'approbation.

Annexe 2 : Méthodologie complète

Cette feuille de route est conçue pour aider les gouvernements à prendre des décisions éclairées sur le moment et la manière de partager des données avec des acteurs externes pour l'innovation en IA. Pour ce faire, nous avons mené une revue de la littérature, élaboré un premier projet de feuille de route, qui a été examiné par le GPAI et des experts externes, testé la feuille de route avec trois partenaires pilotes, incorporé les leçons apprises dans une version révisée, puis organisé une deuxième série d'examen par des experts externes pour la version mise à jour.

Revue de la Littérature Pertinente

Dans le cadre d'une deuxième phase du projet du GPAI, « Le rôle du gouvernement en tant que fournisseur de données pour l'IA », un accent particulier a été mis sur le rapport 2023 du GPAI sur Le rôle du gouvernement en tant que fournisseur de données pour l'intelligence artificielle (GPAI, 2023). Nous avons également examiné le travail plus large du Groupe de travail du GPAI sur le partage de données et les publications pertinentes de ses autres groupes de travail. De plus, nous avons situé notre travail dans le paysage de la recherche plus large sur le partage de données pour l'IA, y compris la littérature académique et grise. Les principales conclusions de ces sources ont été enregistrées pour éclairer la conception du projet initial.

Conception du Projet de Feuille de Route

La feuille de route a été conçue grâce à un processus itératif, commençant par des ateliers et des discussions internes, qui se sont appuyés sur les enseignements de projets antérieurs d'Oxford Insights, comme la [conception d'un projet pilote de partage de données en Colombie](#) (MinTIC, 2022), et sur la revue de la littérature. Sur la base de ces ressources, nous avons développé une feuille de route préliminaire qui a été examinée par un groupe externe d'experts en gouvernance des données ainsi que par le Groupe consultatif de projet (PAG) du GPAI.

Sur la base des commentaires que nous avons reçus des experts et du PAG, nous avons itéré la feuille de route pour créer un premier projet à tester de manière pratique.

Test du projet de feuille de route

Pour nous assurer que notre feuille de route est centrée sur l'utilisateur et qu'elle aborde les principaux facilitateurs, défis et obstacles dans le parcours de partage de données pour l'IA, nous l'avons testée par le biais de projets pilotes avec trois équipes gouvernementales. Nous avons invité les gouvernements du monde entier à manifester leur intérêt et à soumettre une candidature de partenaire, en évaluant chacun à l'aide d'un cadre de notation. Ce cadre a évalué des aspects clés tels que :

- La préparation et l'engagement de l'équipe envers le projet
- Les capacités internes
- Une compréhension claire des objectifs du projet

Un facteur critique pour garantir l'efficacité de la feuille de route était son applicabilité dans différents contextes. Pour ce faire, nous avons soigneusement sélectionné des partenaires pilotes pour représenter un éventail diversifié de :

- Répartition géographique.
- Niveaux de gouvernement.

- Développement économique.
- Maturité de la politique et de la législation sur les données.
- Expérience en matière de partage de données.
- Progrès antérieurs sur la sélection d'un secteur et/ou d'un cas d'usage - visant à avoir un mélange qui nous permet de tester différentes parties de la feuille de route.

Après le processus d'évaluation, nous avons établi des partenariats avec les équipes gouvernementales suivantes :

- Agence pour le gouvernement électronique et l'information (AGESIC) d'Uruguay (niveau national).
- Bureau de la transformation numérique de la Présidence de la Türkiye (niveau national).
- Gouvernement de l'État de Jigawa (niveau de l'État).

Le partenariat impliquait de soutenir les partenaires pilotes dans la mise en œuvre du projet de feuille de route par le biais d'une série d'ateliers et de sessions individuelles de résolution de problèmes adaptés aux besoins spécifiques de chaque équipe gouvernementale. De plus, cette collaboration nous a permis de recueillir des informations précieuses sur la convivialité de la feuille de route. Nous y sommes parvenus en mettant en œuvre des formulaires formels de collecte de commentaires à la fin de chaque étape de décision du projet de feuille de route. De plus, nous avons recueilli d'autres informations sur la convivialité de la feuille de route en observant de près les progrès de chaque pilote.

En plus des projets pilotes, nous nous sommes engagés avec d'autres équipes gouvernementales dans différentes régions pour recueillir des commentaires sur la feuille de route par le biais de sessions individuelles et d'un webinaire organisé en septembre 2024.

Itération de la feuille de route

Les projets pilotes et l'engagement avec les équipes gouvernementales nous ont fourni des informations exploitables pour améliorer la feuille de route, ce qui a conduit au développement d'une version finale itérée. Un certain nombre des principaux enseignements que nous avons tirés des partenariats pilotes comprenaient :

Structure et parcours utilisateur

Notre projet de feuille de route conseillait un chemin linéaire à travers le projet. La feuille de route fournissait une liste élargie de points de décision et conseillait de prendre chaque point de décision à tour de rôle. Dans nos projets pilotes, nous avons constaté que les partenaires avaient des points d'entrée et des séquences de prise de décision préférés différents. Par exemple, la Türkiye a commencé avec un mécanisme et a choisi les cas d'usage en fonction de celui-ci, tandis que l'Uruguay et Jigawa ont tous deux commencé avec des secteurs dans lesquels ils voulaient développer des initiatives. Nous avons mis à jour la structure de la feuille de route pour refléter ces différents parcours utilisateur.

Portée de la feuille de route

Nous avons appris les opportunités et les limites de la manière dont la feuille de route peut fournir un véritable soutien aux utilisateurs. La feuille de route peut aider à structurer les décisions, à fournir des considérations, des ressources et des exemples internationaux. La feuille de route ne peut pas résoudre les défis qui résultent de contextes gouvernementaux uniques. Nous avons mis à jour la feuille de route pour exclure les dépendances à des



méthodes de travail spécifiques ou à des contextes juridiques ou institutionnels. Nous avons inclus des exemples lorsque cela était possible pour illustrer comment les gouvernements travaillent dans les contraintes de leurs contextes spécifiques. Nous proposons des activités et des ressources suggérées lorsque cela est possible pour aider les équipes à surmonter elles-mêmes les défis.

Encadrement des décisions et de la liste de contrôle

L'encadrement des points de décision et la raison pour laquelle ils sont importants sont devenus plus clairs à mesure que nous développons des documents d'atelier. Par exemple, nous avons constaté que les mécanismes de partage de données étaient mieux communiqués comme étant, à leur cœur, des mécanismes de prise de décision adoptés par un ensemble d'acteurs pour la gestion de l'accès et de l'utilisation des données, les exigences d'infrastructure de données et d'infrastructure juridique découlant de ce mécanisme convenu. Nous avons mis à jour l'encadrement de la liste de contrôle et des points de décision de la feuille de route sur la base de cette expérience.

Réponse aux défis auxquels les utilisateurs sont confrontés

Les partenaires pilotes ont rencontré et résolu plusieurs défis au cours des projets pilotes. Ceux-ci incluent une expérience limitée dans l'application des cadres juridiques, la clarification et l'attribution des rôles entre les organisations, l'apprentissage de la manière d'appliquer les nouvelles techniques de technologie de préservation de la confidentialité, la coordination avec les parties prenantes externes. Nous avons inclus dans la feuille de route les documents que nous avons élaborés au cours des projets pilotes pour aider les partenaires à surmonter ces défis, ainsi que des exemples internationaux de la manière dont d'autres gouvernements répondent à ces défis.

La nouvelle version a été examinée par le Groupe consultatif de projet du GPAI ainsi que par nos conseillers experts externes pour garantir un niveau élevé d'assurance qualité. Ces examens ont été cruciaux pour affiner la feuille de route, en veillant à ce qu'elle réponde aux divers besoins des utilisateurs potentiels et qu'elle soit à la fois pratique et alignée sur les dernières meilleures pratiques en matière de partage de données pour l'innovation en IA.



Bibliographie

- Ada Lovelace Institute, (2021a), Legal mechanisms for data stewardship.
www.adalovelaceinstitute.org/wp-content/uploads/2021/03/Legal-mechanisms-for-data-stewardship_report_Ada_AI-Council-2.pdf
- Ada Lovelace Institute, (2021b), Participatory data stewardship: A framework for involving people in the use of data.
https://www.adalovelaceinstitute.org/wp-content/uploads/2021/11/ADA_Participatory-Data-Stewardship.pdf
- ALI-ELI, (2021), Principles for a Data Economy: Data Transactions and Data Rights.
www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ALI-ELI_Principles_for_a_Data_Economy_Final_Council_Draft.pdf
- All of us, (n.d.), National Institutes of Health, www.researchallofus.org/
- DAMA UK, (2013), The Six Primary Dimensions for Data Quality Assessment,
<https://www.sbctc.edu/resources/documents/colleges-staff/commissions-councils/dgc/data-quality-dimensions.pdf>
- Deloitte, (2017), New Technologies Case Study: Data Sharing in Infrastructure A final report for the National Infrastructure Commission,
nic.org.uk/app/uploads/Data-sharing-in-infrastructure.pdf
- Dharyanto, I., Murata, T., Ustriyana, MG., (2023), Indonesian Personal Data Protection Law Update – What to Expect from the Upcoming Implementing Regulation, Data Protection Newsletter,
www.nishimura.com/sites/default/files/newsletters/file/asia_data_protection_231002_en.pdf
- Berryhill, J., Heang, K., Clogher, R., and McBride, K. (2019), Hello, World: Artificial intelligence and its use in the public sector. 10.1787/726fd39d-en.
- Black, G., Stevens, L., (2013), Enhancing Data Protection and Data Processing in the Public Sector: The Critical Role of Proportionality and the Public Interest, SCRIPTed, vol. 10, no. 1, pp. 93-122. <https://doi.org/10.2966/scrip.100113.93>
- C4DC, (n.d), Library, <https://contractsfordatacollaboration.org/library/>
- Ceulemans, H., Boeckx, T., Dillard, V., Mathieu, G., Oberhuber, M. (2021), From competition to collaboration: How secure data sharing can enable innovation, World Economic Forum,
www.weforum.org/agenda/2021/06/collaboration-data-sharing-enable-innovation/
- Chen, Y., Clayton, E. W., Novak, L. L., Anders, S., & Malin, B., (2023), Human-Centered Design to Address Biases in Artificial Intelligence, Journal of medical Internet research, 25, e43251. <https://doi.org/10.2196/43251>
- CNIL, (2024), AI how-to sheets, <https://www.cnil.fr/fr/ai-how-to-sheets>



Copernicus, (n.d.), Copernicus in detail, <https://www.copernicus.eu/>

Data Spaces Support Centre, (2024), Contractual framework, dssc.eu/space/BVE/357074870/Contractual+framework#Data-space-agreements

Daten Ethik Kommission, (2019), Opinion of the Data Ethics Commission, https://www.bfdi.bund.de/SharedDocs/Downloads/EN/Datenschutz/Data-Ethics-Commission_Opinion.pdf?__blob=publicationFile&v=1.

Digital Economy Act, Section 71, (2017), www.legislation.gov.uk/ukpga/2017/30/contents

Digital Future Society (2022), Towards meaningful oversight of automated decision-making systems, digitalfuturesociety.com/app/uploads/2022/11/Towards_meaningful_oversight_of_automated_decision_making_systems.pdf

Doteveryone (2021), 'Consequence Scanning – an agile practice for responsible innovators', doteveryone.org.uk/project/consequence-scanning/

European Commission, (2024), AI Act, digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Commission, (n.d), Are there restrictions on the use of automated decision-making?, commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/dealing-citizens/are-there-restrictions-use-automated-decision-making_en

ESIP Data Readiness Cluster, (2022), Checklist to Examine AI-readiness for Open Environmental Datasets. ESIP. Online resource. <https://doi.org/10.6084/m9.figshare.19983722.v1>

Frontier Economics, (2021), The economic impact of trust in data ecosystems – Frontier Economics for the ODI, theodi.org/insights/reports/the-economic-impact-of-trust-in-data-ecosystems-frontier-economics-for-the-odi-report/

Government of Canada, (2023), Get authority to collect the personal information, The Digital Privacy Playbook, www.canada.ca/en/government/system/digital-government/digital-privacy-playbook/ensure-authority-collect.html

GPAI, (2020), Data Governance Working Group A Framework Paper for GPAI's work on Data Governance, gpai.ai/projects/data-governance/gpai-data-governance-work-framework-paper.pdf

GPAI, (2022a), Data Justice Policy Brief: Putting data justice into practice, gpai.ai/projects/data-governance/data-justice-policy-brief-putting-data-justice-into-practice.pdf

GPAI, (2022b) Enabling Data Sharing for Social Benefit Through Data Trusts: Data Trusts in Climate, Report, March 2022, Global Partnership on AI,



gpai.ai/projects/data-governance/enabling-data-sharing-for-social-benefit-through-data-trusts-in-climate.pdf

GPAI, (2023a), The Role of Government as a Provider of Data for Artificial Intelligence: Phase One Project Report, africa.ai4d.ai/wp-content/uploads/2024/01/DG08-The-Role-of-Government-as-a-Provider-of-Data-for-Artificial-Intelligence-Interim-Report.pdf

GPAI, (2023b), Trustworthy Data Institutional Framework - A practical tool to improve trustworthiness in data ecosystems, <https://gpai.ai/projects/data-governance/DG10%20-%20Trustworthy%20Data%20Institutional%20Framework%20-%20-%20A%20practical%20tool%20to%20improve%20trustworthiness%20in%20data%20ecosystems.pdf>

Hsieh, P.J., Lai, H.M., (2020), Exploring people's intentions to use the health passbook in self-management: An extension of the technology acceptance and health behavior theoretical perspectives in health literacy, Technological Forecasting and Social Change, Elsevier, vol. 161(C), doi.org/10.1016/j.techfore.2020.120328

INSIGHT, (n.d), Data Trust Advisory Board, www.insight.hdrhub.org/data-trust-advisory-board

Jolliffe, D., Mahler, D.G., Veerappan, M., Kilic, K., Wollburg, P., What Makes Public Sector Data Valuable for Development?, 2023. The World Bank Research Observer, Volume 38, Issue 2, August 2023, Pages 325–346, doi.org/10.1093/wbro/lkad004

Koops, B. J., (2021), The concept of function creep. Law, Innovation and Technology, 13(1), 29–56. <https://doi.org/10.1080/17579961.2021.1898299>

Land Transport Authority, (n.d), www.lta.gov.sg/content/ltagov/en.html

MinTIC, (2022), Pasos para pilotear proyectos de intercambios de datos Data Trust, Data Commons y Data Marketplace, https://gobiernodigital.mintic.gov.co/692/articles-238512_recurso_2.pdf

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, ID., Gebru, T., (2019), Model Cards for Model Reporting, In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19). Association for Computing Machinery, New York, NY, USA, 220–229, <https://doi.org/10.1145/3287560.3287596>

Mobility Dataspace, (n.d), mobility-dataspace.eu/

NHS England, (2023), Value Sharing Framework for NHS data partnerships. <https://transform.england.nhs.uk/key-tools-and-info/centre-improving-data-collaboration/value-sharing-framework-for-nhs-data-partnerships/>

NHS, (n.d), Artificial data pilot, digital.nhs.uk/services/artificial-data

OECD, (2024), AI, Data Governance and Privacy: Synergies and Areas of International Co-operation,



www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf

Office for National Statistics, (2023), How we're providing better data to power decision making - the technology behind making critical data assets available on IDS, Blog: ONS Digital, Data and Technology,

onsdigital.blog.gov.uk/2023/01/18/how-were-providing-better-data-to-power-decision-making-the-technology-behind-making-critical-data-assets-available-on-ids/

Open Data Charter, (2023), Finding The 'Rights' Balance: How access to public information and open data complement each other,

medium.com/opendatacharter/finding-the-rights-balance-4a1444823e58

Open Data Institute (ODI), (2019), Data Ecosystem Mapping Tool.

theodi.org/insights/tools/data-ecosystem-mapping-tool/

Open Data Institute (ODI), (2021), 'The Data Ethics Canvas',

theodi.org/article/the-data-ethics-canvas-2021/

Open Data Institute (ODI), (2022), Mapping data ecosystems: methodology.

theodi.org/insights/reports/mapping-data-ecosystems/

Oxford Insights, (2023), Trustworthy AI Self Assessment,

oxfordinsights.com/ai-readiness/trustworthy-ai-self-assessment/

Postigo, A., (2023), Regulation of digital health and health data in the Asia Pacific region, ARTNeT Working Paper Series, No. 230, Asia-Pacific Research and Training Network on Trade (ARTNeT), Bangkok, hdl.handle.net/20.500.12870/6645

Ravi, N., Chaturvedi, P., Huerta, E.A., Liu, Z., Chard, R., Scourtas, A., Schmidt, K.J., Chard, K., Blaiszik, B., Foster, I., (2022), FAIR principles for AI models with a practical application for accelerated high energy diffraction microscopy. Sci Data 9, 657 (2022).

<https://doi.org/10.1038/s41597-022-01712-9>

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, Article 6,

data.europa.eu/eli/reg/2016/679/oj

Republic Act No 10173, (2012), Data Privacy Act of 2012,

<https://privacy.gov.ph/data-privacy-act/#w7>

Republic of Estonia, (2024), Consent Service, Information System Authority,

www.ria.ee/en/state-information-system/people-centred-data-exchange/consent-service

Statistics Law of Turkey, (2005), Article 14,

unstats.un.org/unsd/dnss/docViewer.aspx?docID=2149#:~:text=ARTICLE%201%2D%20The%20purpose%20of,and%20information%2C%20produce%2C%20publish%20and



- Szarfman, A., Levine, J. G., Tanning, J. M., Weichold, F., Bloom, J. C., Soreth, J. M., Geanacopoulos, M., Callahan, L., Spotnitz, M., Ryan, Q., Pease-Fye, M., Brownstein, J. S., Ed Hammond, W., Reich, C., & Altman, R. B., (2022), Recommendations for achieving interoperable and shareable medical data in the USA, Communications medicine, 2, 86. <https://doi.org/10.1038/s43856-022-00148-x>
- Tony Blair Institute for Global Change, (2022), Reaping the Rewards of the Next Technological Revolution: How Africa Can Accelerate AI Adoption Today, www.institute.global/insights/tech-and-digitalisation/reaping-rewards-next-technological-r-evolution-how-africa-can-accelerate-ai-adoption-today
- UK Data Service, (n.d), Apply to access non-ONS controlled data ukdataservice.ac.uk/find-data/access-conditions/secure-application-requirements/apply-to-access-non-ons-data/
- UNESCO, (2023), AI Ethics Impact Assessment, unesdoc.unesco.org/ark:/48223/pf0000386276/PDF/386276eng.pdf.multi
- UNESCO, OECD, (2024), G7 toolkit for artificial intelligence in the public sector; report prepared for the 2024 Italian G7 presidency and the G7 digital and tech working group, unesdoc.unesco.org/ark:/48223/pf0000391566
- The White House, (2022), OSTP Issues Guidance to Make Federally Funded Research Freely Available Without Delay, www.whitehouse.gov/ostp/news-updates/2022/08/25/ostp-issues-guidance-to-make-federally-funded-research-freely-available-without-delay/
- World Bank, (2019), ID4D Practitioner's Guide: Version 1.0 (October 2019), Washington, DC: World Bank, id4d.worldbank.org/guide/data-protection-and-privacy-laws
- World Economic Forum, (2020), AI Procurement in a Box: AI Government Procurement Guidelines, www.nist.gov/system/files/documents/2021/08/23/ai-rmf-rfi-0039-1.pdf
- Yates D., Beale T., Marshall S., Parr M., (2018), Designing data sharing agreements: a checklist. In Gates Open Res 2018. <https://doi.org/10.21955/gatesopenres.1114886.1>
- Young, A., Verhulst, S., (2017), Aclímate Colombia: Open Data to Improve Agricultural Resiliency, Open Data's Impact, odimpact.org/case-aclimate-colombia.html
- Ziesche, S. (2023), Open data for AI: What now?, UNESCO, doi.org/10.58338/OGYU7382