

Groupe de travail chargé de la gouvernance des données

Novembre 2020 - Sommet de Montréal sur le Partenariat
mondial sur l'intelligence artificielle (PMIA)



GPAI / THE GLOBAL PARTNERSHIP
ON ARTIFICIAL INTELLIGENCE

Le présent rapport a été élaboré par des experts du groupe de travail chargé de la gouvernance des données du PMIA. Le rapport reflète les opinions personnelles des experts du PMIA et ne reflète pas nécessairement le point de vue des organisations des experts, du PMIA, de l'OCDE ou de leurs membres respectifs.



Mot de bienvenue des coprésidentes	4
i) Présentation du groupe de travail	6
Membres du Groupe de travail chargé de la gouvernance des données du PMIA	7
ii) Notre mandat	8
iii) Méthodes de travail	8
Calendrier du groupe de travail	9
iv) Recommandations préliminaires et résultats pour le Sommet	9
v) Priorités pour le premier semestre 2021	11
vi) Vision à long terme	16
Annexe 1	18
Annexe 2	20



Dr. Jeni Tennison

Vice-présidente et conseillère stratégique en chef
Open Data Institute



Dr. Maja Bogataj Jančič

Fondatrice et directrice
Intellectual Property Institute

Une bonne gouvernance des données, recueillies, utilisées et partagées de manière responsable et fiable, constitue l'un des défis fondamentaux du développement de l'intelligence artificielle (IA). La gouvernance des données doit être compatible avec les droits de la personne, l'inclusion, la diversité, l'innovation, la croissance économique et les avantages pour la société, en cohérence avec les objectifs de développement durable de l'Organisation des Nations Unies (ONU).

En 2020, l'importance que revêt la bonne gouvernance des données a été démontrée et mise à l'épreuve. De l'efficacité des systèmes de recherche des contacts à la sécurisation de l'accès aux données pertinentes sur les données chimiques et sur les médicaments pour la découverte et le repositionnement de médicaments assistés par l'IA, l'année 2020 a mis au défi ceux qui travaillent dans la gouvernance des données, et nous tous, bien sûr, au niveau personnel. Il y a eu cette année des lueurs d'espoir de notre côté : nous avons mis sur pied ce groupe de travail; nous avons également appris à connaître ce groupe d'experts : un groupe énergique, passionné et collaboratif qui a déployé tant d'énergie et donné une perspective mondiale à cette initiative.

Nous sommes extrêmement reconnaissants de l'engagement dont ils ont fait preuve. Collectivement, nous partageons une même conviction forte de la valeur et de l'importance que revêt la mission du PMIA et nous sommes animés du désir de voir son incroyable potentiel se réaliser. Il est probable que la gouvernance des données soit un élément fondamental dans un large éventail de projets du PMIA et nous avons la chance de disposer d'un groupe d'experts prêt à apporter leur soutien. Notre mandat reflète étroitement et délibérément celui de la mission globale du PMIA selon laquelle ce groupe d'experts doit « rassembler des preuves, façonner la recherche, entreprendre des projets d'IA appliquée et fournir son expertise sur la gouvernance des données, afin de promouvoir la collecte des données pour l'IA¹, leur utilisation, partage, archivage et suppression de manière compatible avec les droits de la personne, l'inclusion, la diversité, l'innovation, la croissance économique et les avantages pour la société, tout en cherchant à atteindre les objectifs de développement durable des Nations Unies ».

Le présent rapport décrit notre travail préliminaire au cours des six premiers mois : l'élaboration d'un cadre qui établit la portée et les modalités du groupe de travail et une enquête que nous avons commandée sur le rôle des données dans le contexte de l'IA ainsi que quelques recommandations initiales sur la poursuite de la collaboration internationale; il décrit également la manière dont nous avons l'intention de travailler avec d'autres groupes de travail et de concentrer nos efforts sur le soutien des projets axés sur l'action qui contribueront à nous acquitter de notre mandat. Nous avons hâte de discuter du prochain chapitre des travaux du PMIA.

¹À cette fin, le mandat s'appuie sur les définitions figurant dans le document : [Recommandations de l'OCDE sur l'intelligence artificielle](#).

(i) Présentation du groupe de travail

Notre groupe de travail est composé de 27 experts, dont deux observateurs, provenant de 17 pays ayant l'expérience des aspects techniques, juridiques et institutionnels de la gouvernance des données. Fidèles à l'ambition mondiale du PMIA, ils combinent des connaissances provenant de plusieurs secteurs, comme la communauté scientifique, l'industrie, la société civile et les organisations internationales. Ils apportent au groupe des points de vue différents allant de ceux de concepteurs d'ensembles de données de la COVID-19 à ceux d'intervenants dans les collectivités autochtones en vue d'améliorer leur représentation. Des discussions fructueuses se sont engagées à l'occasion des délibérations du groupe de travail; nous sommes chanceux d'avoir un groupe très énergique, passionné et collaboratif apportant de nombreux points de vue qui ne se limitent pas à un seul pays.

Nous souhaitons nous appuyer sur ces discussions et, en particulier, étoffer la diversité au groupe. Par conséquent, nous sommes impatients d'accroître le nombre de membres du PMIA en 2021 et d'accueillir des experts de ces nouveaux membres.



Membres du Groupe de travail chargé de la gouvernance des données du PMIA

Membres du groupe de travail

Jeni Tennison (coprésidente) – Open Data Institute (Royaume-Uni)
Maja Bogataj Jančič (coprésidente) – Intellectual Property Institute (Slovénie)
Alejandro Pisanty Baruch – National Autonomous University (Mexique)
Alison Gillwald – Research ICT Africa (Afrique du Sud/UNESCO)
Bertrand Monthubert – Occitanie Data (France)
Carlo Casonato – University of Trento (Italie)
Carole Piovesan – INQ Data Law (Canada)
Christiane Wendehorst – European Law Institute/University of Vienna (UE)
Dewey Murdick – Center for Security and Emerging Technology (États-Unis)
Hiroshi Mano – Data Trading Alliance (Japon)
Iris Plöger – Federation of German Industries (Allemagne)
Jeremy Achin – DataRobot (États-Unis)
Josef Drexler – Max Planck Institute (Allemagne)
Kim McGrail – University of British Columbia (Canada)
Matija Damjan – University of Ljubljana (Slovénie)
Neil Lawrence – University of Cambridge (Royaume-Uni)
Nicolas Mialhe – The Future Society (France)
Oreste Pollicino – University of Bocconi (Italie)
Paola Villerreal – National Council for Science and Technology (Mexique)
Paul Dalby – Australian Institute of Machine Learning (Australie)
P. J. Narayanan – International Institute of Technology, Hyderabad (Inde)
Shameek Kundu – Standard Chartered Bank (Singapour)
Takashi Kai – Hitachi (Japon)
Teki Akuetteh Falconer – Africa Digital Rights Hub (Ghana/UNESCO)
Te Taka Keegan – University of Waikato (Nouvelle-Zélande)
V. Kamakoti – International Institute of Technology, Madras (Inde)
Yeong Zee Kin – Infocomm Media Development Authority (Singapour)

Observateurs

Elettra Ronchi – OECD
Jaco Du Toit – UNESCO



ii) Notre mandat

En tant que groupe, notre mandat est étroitement aligné sur la mission globale du PMIA. Notre groupe de travail a pour objectif de « rassembler des preuves, de façonner la recherche, d'entreprendre des projets d'IA appliquée et de fournir son expertise sur la gouvernance des données, afin de promouvoir la collecte des données pour l'IA, leur utilisation, partage, archivage et suppression de manière compatible avec les droits de la personne, l'inclusion, la diversité, l'innovation, la croissance économique et les avantages pour la société, tout en cherchant à atteindre les objectifs de développement durable des Nations Unies ».

Il existe de toute évidence des interactions entre la gouvernance des données et les domaines de compétences des autres groupes de travail, en particulier ceux chargés de l'IA responsable et de la commercialisation et de l'innovation et nous avons donc l'intention de travailler de concert avec eux dans les domaines qui se recoupent. Le groupe de travail a également la possibilité d'aider à coordonner les ambitions du PMIA en matière d'IA appliquée, de façonner les projets réalisés ou financés par les membres du PMIA ainsi que dans le cadre de ses partenariats élargis, et d'influencer les recommandations politiques formulées par l'OCDE au moyen de ses travaux. Plus largement, nous cherchons également à ce que les personnes chargées de rechercher, de réfléchir et de mettre en œuvre des pratiques de gouvernance des données en matière d'IA trouvent nos travaux utiles.

iii) Méthodes de travail

Dans notre façon d'aborder notre travail, nous nous sommes engagés très tôt et [publiquement](#) à placer les valeurs d'ouverture, de transparence, de collaboration et de diversité au cœur de notre processus. Comme le montre notre calendrier ci-après, ces valeurs ont façonné très concrètement nos travaux et nous continuerons à accroître notre diversité et à renforcer notre collaboration en 2021. Nous nous réjouissons en particulier à l'idée de profiter de nouveaux partenariats, de développer des collaborations entre les groupes de travail et d'accueillir des experts provenant de nouveaux pays au fur et à mesure de l'élargissement du PMIA.



Calendrier du groupe de travail

JUILLET

- **Présentation des coprésidentes** (2 juillet).
- **Première réunion du groupe de travail** (24 juillet) : Présentations, accord sur le mandat et livrables du Sommet (*le cadre de gouvernance des données et le Rôle des données dans le contexte de l'IA*).
- **Documents de travail** sur les approches de gouvernance des données recueillis auprès des experts du groupe de travail.

AOÛT

- **Mise sur pied des comités directeurs de projet** : Réunions hebdomadaires à partir de la semaine du 23 août.
- **Christiane Wendehorst désignée comme chef de projet** du cadre de gouvernance des données.
- **Blogue de présentation du groupe de travail publié auprès de l'OCDE, comprenant une demande de propositions** et un mandat publié sur *le Rôle des données dans le contexte de l'IA*.

SEPTEMBRE

- **Deuxième réunion du groupe de travail** (2 septembre) : Kim McGrail est désignée comme la liaison avec le groupe de travail sur la pandémie; séances en petits groupes sur le cadre de gouvernance des données (sur le cycle de vie des données, l'écosystème des données/modèle d'acteur et les modèles de cadre de gouvernance des données).
- **Premier cycle d'évaluation des propositions sur le Rôle des données dans le contexte de l'IA** par un groupe d'experts du groupe de travail chargés de l'évaluation (à partir de la semaine du 7 septembre).
- **Deuxième cycle d'évaluation des propositions et désignation du Digital Curation Centre, de la School of Informatics de Edinburgh University et du Trilateral Research** comme consultants sur le Rôle des données dans le contexte de l'IA (à partir de la semaine du 14 septembre).
- **Première ébauche du cadre de gouvernance des données communiquée** au sein du groupe de travail et lancement du sondage (22 septembre).
- **Troisième réunion du groupe de travail** (28 septembre) : Présentation de l'équipe de projet chargée du Rôle des données dans le contexte de l'IA et présentation de la première ébauche du cadre de gouvernance des données accompagnée d'une discussion en séance plénière sur les questions du sondage, suivies par des discussions en petits groupes sur la feuille de route du cadre de gouvernance des données; discussion sur la sensibilisation.

OCTOBRE

- **Établissement de la liste d'experts des pays du Sud** comportant les recommandations des membres du groupe de travail à des fins d'invitations à des événements ouverts.
- **Réunion de tous les coprésidents** pour comparer les progrès réalisés et discuter des synergies potentielles (16 octobre).
- **Premier atelier sur le Rôle des données dans le contexte de l'IA** composé de participants et d'intervenants invités supplémentaires et consacré au thème du « Rôle des données dans le développement technologique du langage humain pour les langues en manque de ressources » (19 octobre).
- **Quatrième réunion du groupe de travail** : Présentation de la version « bêta » du cadre de gouvernance des données et première ébauche; discussions sur le Rôle des données dans le contexte de l'IA, séances en petits groupes sur la disponibilité/l'accessibilité des données pour l'IA; répercussions sur les plans socio-économique, environnemental et juridique; et notre public cible pour les recommandations (28 octobre).

NOVEMBRE

- **Deuxième atelier sur le Rôle des données dans le contexte de l'IA** composé de participants et d'intervenants invités supplémentaires et consacré au thème des systèmes juridiques fondés sur les données en examinant les conséquences sur les plans social, juridique et éthique (2 novembre).
- **Version « bêta » du cadre de gouvernance des données** publiée sur le blogue AI Wonk de l'OCDE; une période de deux semaines est prévue pour que la communauté élargie puisse ajouter des commentaires (3 novembre).
- **Cinquième réunion du groupe de travail** (19 novembre) : Courte présentation de la deuxième ébauche sur le Rôle des données dans le contexte de l'IA; séance « improvisée » du conseil sur les plans du premier semestre 2021 (élaboration d'une longue liste de notes conceptuelles en phase avec la feuille de route du cadre).
- **Dernier atelier sur le Rôle des données dans le contexte de l'AI** portant sur les sources de données responsables et fiables, ouvertes et FAIR pour l'IA (23 novembre).
- **Réunion de tous les coprésidents** avant le Sommet (27 novembre).

DÉCEMBRE

- **Présentation des résultats finaux et atelier ouvert sur les prochains projets lors du Sommet.**

iv) Recommandations préliminaires et résultats pour le Sommet

Dans le cadre des préparatifs du Sommet, il a été convenu de mettre en avant les deux principaux résultats suivants :

1. **un cadre pour le travail du PMIA sur la gouvernance des données** : Poser les jalons pour tous les futurs projets du groupe de travail, donnant une vue d'ensemble des modalités les plus pertinentes et définissant la compréhension de la gouvernance des données dans le contexte de l'AI du groupe de travail; et
2. **une enquête sur le Rôle des données dans le contexte de l'IA** : Afin de compléter et d'approfondir les thèmes du cadre, cette enquête détermine l'importance des données pour le développement de l'IA et cerne à la fois les domaines dans lesquels plus de données seraient utiles comme des ensembles de données spécifiques, ouverts, qui pourraient mériter un soutien national ou une collaboration internationale et les domaines dans lesquels des préjudices peuvent découler de l'accès aux données, de leur collecte ou de leur utilisation.

Christiane Wendehorst, professeure de droit civil à l'University of Vienna et présidente de l'European Law Institute a assumé la direction du cadre. Elle a été soutenue par deux assistants Nina Thomic et Yannic Duller. Comme le montre la section iii, le cadre a été élaboré avec l'entière collaboration du groupe de travail élargi pendant le déroulement de nombreux ateliers et dans le contexte de sondages et d'ébauches organisés au cours des derniers mois.

Au plus haut niveau, le cadre aborde quatre domaines :

1. **Le rôle des données dans le contexte de l'IA** : notamment les données dans le développement et le déploiement de l'IA ainsi que le cycle de vie des données.
2. **Les raisons justifiant l'importance de la gouvernance des données** : notamment les études de cas illustrant le besoin d'une bonne gouvernance des données, le rôle et les responsabilités des différents acteurs et les principes de la gouvernance des données.
3. **Les paramètres de la gouvernance des données** : notamment les catégories de données, les écosystèmes de données et les droits concernant les données.
4. **Une feuille de route pour les travaux futurs du groupe de travail décrivant la manière dont celui-ci mettra l'accent sur trois types d'approches de gouvernance des données** : 1) les approches techniques (p. ex., les technologies permettant d'améliorer le respect de la vie privée, les techniques de détection et de correction des partis pris), 2) les approches juridiques (p. ex., la prise en considération du droit de la propriété intellectuelle, du droit en matière de protection des données) et 3) les approches organisationnelles/institutionnelles (p. ex., les représentants ou fiduciaires de données, espaces communs de données).

Il convient de noter que le cadre doit être considéré comme un document « à caractère évolutif » qui se transformera en fonction des besoins afin de suivre les évolutions dans ce domaine.

En ce qui concerne l'enquête liée au rôle des données dans le contexte de l'IA, nous avons fait appel à [un consortium](#) dirigé par l'University of Edinburgh, qui combine d'importantes compétences techniques et juridiques, composé de la School of Informatics, du Digital Curation Centre et du Trilateral Research. Il creuse encore plus en profondeur les questions soulevées au sein du cadre et identifie les domaines dans lesquels le PMIA pourrait avoir des répercussions sur le renforcement de la collaboration internationale. Il aborde les domaines suivants :

- le développement de l'IA et le rôle des données à chaque étape;
- les types de données utilisées pour le développement de l'IA;
- les caractéristiques des données qui influencent le processus ou le résultat du développement de l'IA;
- les répercussions sur les plans socio-éthique, économique et environnemental des données dans le contexte de l'IA;
- la loi et la transparence en tant que modificateurs de l'impact des données dans le contexte de l'IA;
- la disponibilité et l'accessibilité des données pour le développement de l'IA : qualité des données et défis dans trois domaines (réponse à la pandémie, technologies du langage humain pour les langues en manque de ressources et applications de l'IA dans le système judiciaire pénal); et
- les recommandations sur les points où le PMIA pourrait améliorer la collaboration internationale sur la gouvernance des données.

Nous comptons publier *le Rôle des données dans le contexte de l'IA* immédiatement après le Sommet.

Le comité directeur composé d'experts du groupe de travail dont la liste est établie dans l'annexe 2 pilotera les deux projets.

Recommandations préliminaires

Le cadre établit une portée, une structure et comporte un vocabulaire ayant fait l'objet d'un accord pour les travaux du groupe de travail. Nous développons une compréhension commune de la gouvernance des données dans le contexte de l'IA (y compris l'ensemble des termes et des cadres conceptuels convenus qui seront utilisés) et introduisons une feuille de route sur la façon dont le PMIA devrait structurer son travail de gouvernance des données.

La feuille de route décrit de quelle manière le groupe de travail mettra l'accent sur trois types d'approches de gouvernance des données :

1. **Les approches techniques** (p. ex., les technologies permettant d'améliorer le respect de la vie privée, les techniques de détection et de correction des partis pris).
2. **Les approches juridiques** (p. ex., la prise en considération du droit de la propriété intellectuelle, du droit en matière de protection des données).
3. **Les approches organisationnelles/institutionnelles** (p. ex., les représentants ou fiduciaires de données, espaces communs de données).

En fixant cette orientation, la feuille de route adopte une vision horizontale de l'approche du groupe de travail. Cette vision reflète la nature fondamentale de la gouvernance des données et est adaptée à notre expertise, tout en maintenant la flexibilité et une utilisation plus large de notre travail.

L'enquête sur le « *Rôle des données dans le contexte l'IA* » comprend quatre recommandations qui pourraient améliorer la coopération internationale sur la gouvernance des données, en fournissant des orientations davantage axées sur les projets qui complètent la feuille de route et qui guideront nos prochaines étapes en tant que groupe de travail. Pour chacune des recommandations, elle se réfère à des initiatives actuelles sur lesquelles le PMIA pourrait s'appuyer et qu'il pourrait soutenir.

Recommandation 1 : Le groupe de travail chargé de la gouvernance des données devrait œuvrer à façonner les normes et les pratiques exemplaires pour la gouvernance des données en vue de faciliter l'accès à des données de bonne qualité pour les projets et systèmes d'IA. Les mesures concrètes comprennent :

Mesure 1a : Créer des lignes directrices sur la gestion des données pour les projets et systèmes d'IA qui tiennent compte de toutes les étapes du processus de développement de l'IA, de la création et la collecte des données à leur préservation et suppression. Le groupe de travail devrait également se pencher sur la création sur un modèle de plan pour la gestion des données pour les projets et systèmes d'IA qui permettra de saisir les renseignements nécessaires pour étayer la repérabilité, la documentation, la caractérisation, la confiance et la transparence (voir les recommandations 1b à 1e), et ainsi favoriser une réutilisation améliorée et éclairée des données pour l'IA.

Mesure1b : Appuyer les bonnes pratiques concernant le dépôt et le catalogage des sources de données d'IA afin de les rendre plus repérables et accessibles. Ce travail devrait se concentrer sur les points suivants :

- Mener une étude de faisabilité sur différentes options pour améliorer l'accès des données pour les projets et systèmes d'IA. Les options peuvent comprendre la mise en place d'un entrepôt de données d'IA ou d'un catalogue de métadonnées ou la création d'un réseau d'entrepôts existants et d'un point d'accès et de repérage unique.
- Travailler de concert avec des initiatives qui favorisent l'adoption des principes FAIR, de même que le mouvement de science ouverte, et s'assurer que l'IA a des entrées sur tout problème particulièrement pertinent pour des pratiques en matière de données particulières à ce domaine.
- Collaborer avec le groupe de travail sur la pandémie pour la mise en œuvre de leur recommandation d'un portail central sur la réponse aux pandémies. L'expérience tirée de cette collaboration pourra ensuite être reportée et appliquée à d'autres domaines.

Mesure 1c : Élaborer des lignes directrices pour la documentation d'ensembles de données et les métadonnées pour les projets et systèmes d'IA. Ce travail devrait se concentrer sur les points suivants :

- Définir une norme d'information minimale pour la description des sources des données d'IA, basée sur les bonnes pratiques de documentation des données.
- Élaborer des orientations sur la meilleure façon d'intégrer la provenance et le lignage des données dans les métadonnées afin d'améliorer la traçabilité des ensembles de données. Examiner le travail des initiatives dans ce domaine et collaborer sur la définition de normes et de bonnes pratiques à ce sujet.
- Définir la manière dont les problèmes de droits de propriété intellectuelle et de concessions de licences relatives aux données sont présentés dans la documentation.

Mesure1d : Élaborer des lignes directrices et des suggestions sur la documentation de caractérisation des données pour l'harmonisation de chaque projet ou système. Ces lignes directrices comprendraient une orientation sur les points suivants :

- Comment définir un cas d'utilisation de données souhaitées pour le projet ou le système, c'est-à-dire les données nécessaires pour réaliser les objectifs du projet ou système afin de s'assurer que les données sélectionnées sont adéquates.
- Comment identifier les sensibilités de données afin d'inclure les problèmes juridiques et réglementaires liés au cas d'utilisation et veiller à les atténuer.

- Comment évaluer l'exhaustivité des données existantes (pour les réutilisateurs) et veiller à l'exhaustivité des données créées.
- Comment procéder à l'amélioration des données et gérer les données générées par le système d'IA.

Mesure 1e : Élaborer des lignes directrices pour les créateurs de données concernant la transparence pour les utilisateurs de données relativement à la création et au contenu de l'ensemble de données, afin d'accroître la confiance dans ces ressources de données et leur utilisation. Cette recommandation est étroitement liée aux recommandations 1c et 1d, mais ce travail se concentrera plus particulièrement sur la manière de favoriser la confiance des utilisateurs de données dans les ensembles de données qu'ils comptent utiliser dans leurs projets ou systèmes d'IA. Ce travail se concentrera sur les points suivants :

- Représentativité et couverture des données. Indiquer s'il y a des problèmes concernant la représentativité et la couverture dans l'ensemble de données, et le cas échéant, énumérer les mesures qui ont été prises pour éliminer les biais dans les ensembles de données.
- Exactitude et pertinence des données. Indiquer les mesures qui ont été prises pour vérifier l'exactitude des données.
- Définir les problèmes juridiques et éthiques identifiés relativement aux données et la manière dont ils ont été résolus.
- Élaborer des mécanismes fiables (p. ex., des badges de certification) pour illustrer que les ensembles de données ont fait l'objet de processus intégrant les vérifications susmentionnées.

Recommandation 2 : Soutenir la création de sources de données de bonne qualité et accessibles pour combler les lacunes dans les données pour les domaines prioritaires, conformément aux objectifs de développement durable de l'ONU, par l'entremise de recherches ciblées et de collaboration avec les initiatives dans ce domaine.

L'accent devrait être mis sur le soutien de la création de sources de données de bonne qualité et accessibles, en accord avec les pratiques exemplaires de gouvernance des données. Les mesures à suivre pour travailler avec les gouvernements, les initiatives « IA pour le bien social » et les parties prenantes pertinentes, et ainsi soutenir et établir des sources de données fiables dans les domaines prioritaires devraient être décrites. Le groupe de travail devrait se pencher particulièrement sur les domaines où l'investissement est peu probable, et travailler de concert avec d'autres groupes de travail et le PMIA afin que des mesures soient prises pour rendre les données accessibles à l'échelle mondiale. Dans le cadre de ce travail, il est important que l'étude cerne également les lacunes dans la création d'ensembles de données à partir de sources de données disparates pour la compréhension de problèmes complexes. La réponse à la pandémie constitue un exemple pour lequel le manque d'ensembles de données comprenant des données socio-économiques, des données sur les dossiers de santé, et des données génomiques mène à des risques importants pour la santé publique.

Recommandation 3 : Mener des recherches sur la manière d'améliorer le partage transfrontalier de données et rédiger des lignes directrices pour les organisations sur la manière de gérer les obstacles actuels, comme :

- Droits de propriété intellectuelle
- Réglementation en matière de protection des données et de la vie privée
- Souveraineté des données

De plus, le groupe de travail devrait se pencher sur la meilleure façon d'appuyer les développements technologiques, comme les technologies d'apprentissage fédéré et les technologies permettant d'améliorer le respect de la vie privée pour le partage de données, à titre de mesures d'atténuation possibles des défis juridiques, particulièrement dans le cas de données personnelles, et encourager leur développement et adoption lorsque possible.

Recommandation 4 : Mener des recherches ciblées portant sur le vaste sujet de l'injustice des données et des préjudices qui découlent de pratiques en matière de données à l'échelle mondiale et cerner les voies pour contrer les problèmes actuels. Des analyses devraient être réalisées sur les mécanismes possibles qui peuvent surmonter les défis identifiés. Le groupe de travail devrait rechercher des initiatives qui travaillent dans ce domaine et les appuyer dans la création de mécanismes concrets pour redresser les impacts préjudiciables des données dans le contexte de l'IA. Nous proposons les domaines prioritaires suivants :

- La souveraineté des données autochtones et les frictions possibles relativement à la mise en place des principes FAIR et de la liberté d'accès aux données.
- Les partis pris dans les données et les répercussions sur la société et les droits de la personne. L'inclusivité dans les données d'IA afin de réaliser des avantages à une plus grande échelle et d'éviter les préjudices.
- Les préjudices environnementaux émanant du traitement et du stockage des données, et la manière de les atténuer.

Le renforcement des capacités en matière de données dans les pays du Sud, par l'entremise de collaborations et de réseaux internationaux axés sur la construction d'infrastructures matérielles et accessoires dans la région.

Nous tenons à remercier le Digital Curation Centre, la School of Informatics et le Bayes Centre pour ces recommandations, et la contribution qu'ils apportent maintenant en orientant la prochaine phase des travaux du groupe de travail pour le premier semestre 2021.

v) Priorités pour le premier semestre 2021

Le groupe de travail se base maintenant sur ces deux premiers résultats pour élaborer une série de *notes conceptuelles* au premier semestre 2021 pour des projets et des programmes de travail qui pourraient faire progresser la mission du PMIA, et qui pourraient être financés par ses membres en partenariat avec d'autres intervenants. L'objectif est de fournir un moyen de passer à l'action en formulant un ensemble de défis en termes d'activité.

Les notes conceptuelles comprendront :

- une description du problème et le contexte de ce qui est connu jusqu'à présent;
- l'impact du projet attendu à long terme;
- les résultats plus concrets du projet;
- les activités concernées;
- leurs résultats;
- éventuellement, une description des ressources nécessaires au projet

Nous pouvons formuler les défis dans les termes « nous ne savons pas/nous ne connaissons pas... » (recherche) :

- comment faire
- les répercussions de quelque chose
- comment quelque chose fonctionne actuellement

Ou dans les termes « nous n'avons pas... » (développement) :

- la plateforme dont nous avons besoin (par exemple pour la recherche d'ensembles de données)
- l'ensemble de données ou les ensembles de données dont nous avons besoin
- l'ensemble des orientations dont nous avons besoin

Elles s'inspireront des recommandations formulées dans le *Rôle des données dans le contexte l'IA*, seront structurées et conçues comme un ensemble de défis en phase avec le cadre.

Le groupe de travail a profité de sa dernière réunion avant le Sommet pour établir la liste des défis potentiels présélectionnés. Cependant, nous souhaitons continuer à travailler comme nous avons commencé à le faire, c'est-à-dire avec une communauté élargie pour aborder ces défis, et le Sommet offre donc une excellente occasion d'amorcer cette prochaine phase de travail. Le groupe de travail apportera sa liste de présélection comportant ses idées initiales devant être discutées et testées en séance plénière élargie et prévoit de classer les défis par ordre de priorité en appliquant les critères suivants :

- Lesquels contribuent particulièrement à la gouvernance des données dans un contexte de l'IA?
- Lesquels aideraient à progresser vers l'atteinte des objectifs de développement durable, c'est-à-dire à apporter un avantage pour l'intérêt général?
- Lesquels nécessitent une collaboration internationale?
- Lesquels nécessitent une collaboration entre les gouvernements, les entreprises, les universités et le troisième secteur?

Nous avons hâte d'entamer le prochain chapitre de nos travaux.

vi) Vision à long terme

En ce qui concerne sa vision à plus long terme, le groupe de travail sera guidé par la réalisation de la mission globale du PMIA². Les notes conceptuelles élaborées au premier semestre de 2021 par le groupe de travail mentionneront un ensemble de résultats pratiques qui pourraient être atteints avec l'aide du PMIA et préciseront les moyens de les atteindre. Grâce à la collaboration et au partenariat, ces résultats occuperont une place centrale dans les discussions du groupe de travail au cours des deux prochaines années. Ceci marque alors l'ouverture d'un cycle de vie de projet pour les années à venir.

Le groupe de travail s'efforcera également de collaborer avec d'autres groupes de travail, soit directement sur ses propres projets, soit en « détachant » ses experts auprès d'autres groupes de travail pour les conseiller sur les aspects de gouvernance des données de leurs projets d'IA appliquée.

²Telle que définie dans le mandat, la mission du PMIA consiste « à soutenir et à orienter l'adoption responsable de l'IA qui est fondée sur les droits de la personne, l'inclusion, la diversité, l'innovation, la croissance économique et les avantages pour la société, tout en cherchant à atteindre les objectifs de développement durable des Nations unies. Le PMIA facilitera la collaboration internationale multipartite, orientée sur des projets, avec la communauté scientifique, l'industrie, la société civile, les organisations internationales et les pays, en tenant particulièrement compte des intérêts et de la contribution des pays émergents et en voie de développement. Il surveillera aussi le déroulement des travaux réalisés sur le plan national et international et s'en inspirera pour cerner les lacunes, maximiser la coordination et faciliter la collaboration internationale en matière d'IA. »

Le cadre indique que les mandats du groupe de travail chargé de la gouvernance des données et du groupe de travail chargé de l'IA responsable en particulier sont étroitement liés et se recoupent dans une certaine mesure. D'une manière générale, le groupe de travail chargé de l'IA responsable se penchera davantage sur la manière de modéliser le développement de l'IA et d'utiliser les ensembles de données, afin que l'IA soit façonnée et fonctionne de manière responsable (par exemple, sans partis pris excessifs). Par conséquent, le groupe de travail chargé de la gouvernance des données portera son attention en premier lieu sur la façon de collecter et de gérer les données de manière responsable, en particulier en tenant compte de la situation des parties qui sont associées d'une manière ou d'une autre à l'origine et au contexte des données ou qui peuvent être concernées par l'utilisation des données (par exemple les personnes visées et celles appartenant aux collectivités au sujet desquelles les données sont collectées).

Il existe cependant de nombreux autres liens entre les groupes de travail et nous sommes impatients de leur offrir une aide fondamentale dans leurs projets futurs.

vii) ANNEXE 1

Mandat du groupe de travail chargé de la gouvernance des données du PMIA

Portée du mandat du groupe de travail

- Le groupe de travail chargé de la gouvernance des données a pour mission **de rassembler des preuves, de façonner la recherche, d’entreprendre des projets d’IA appliquée et de fournir son expertise** sur la gouvernance des données afin de promouvoir³ la collecte des données pour l’IA, leur utilisation, partage, archivage et suppression de manière compatible avec **les droits de la personne, l’inclusion, la diversité, l’innovation, la croissance économique et les avantages pour la société, tout en cherchant à atteindre les objectifs de développement durable des Nations Unies.**
- Les coprésidents se concerteront avec leurs homologues du **groupe de travail chargé de l’IA** responsable pour s’aligner sur les cas pratiques propres à un secteur et les projets d’IA appliquée, ainsi qu’avec le **groupe de travail chargé de la commercialisation et de l’innovation** sur les questions de propriété intellectuelle. Le **Centre d’expertise de classe mondiale pour l’avancement de l’intelligence artificielle (« CEIMIA »)** travaillera de concert avec l’**OCDE** pour consolider les résultats.

Livrables à présenter lors de la séance plénière du groupe d’experts multipartite

- **Un bref cadre définissant la gouvernance des données⁴**, notamment une analyse documentaire d’autres recherches sur la gouvernance des données, et divisant le sujet en domaines tels que les technologies, les lois/politiques et les organisations/institutions. Il devrait également mettre en lumière la nécessité de traiter à la fois de l’ouverture et de la clôture définitive des données, et la nécessité de parler à la fois des données personnelles, des données non personnelles (y compris les conséquences sur les protections de la propriété intellectuelle) et des données collectives de groupe (p. ex., pour une communauté autochtone comme les Maoris). Il devrait aborder les liens avec les autres groupes de travail (p. ex., toute exception nécessaire en matière de données dans le droit de la propriété intellectuelle et les lois sur le droit d’auteur). L’objectif est d’établir une portée et une structure générales aux travaux du groupe de travail.
- **Une description du rôle des données dans le développement et l’utilisation de l’IA fondée sur des valeurs centrées sur l’humain**, y compris répertorier les ensembles de données particuliers (ou leur type) (p. ex., des ensembles de données de reconnaissance faciale, des ensembles de données soutenant le développement de véhicules autonomes) qui contribuent particulièrement à l’innovation en matière d’IA. Cette recherche devrait procéder à une analyse documentaire sur les avantages et les risques économiques et sociaux qui découlent d’un meilleur accès aux données et de leur réutilisation. Cette recherche a pour objectif de déterminer l’importance des données pour le développement de l’IA et de répertorier des ensembles de données spécifiques (ouverts) qui pourraient mériter un soutien national ou une collaboration internationale. Des lignes directrices sur la façon de rendre ces données ouvertes et réutilisables pour le développement de l’IA (concessions de licences ou exceptions à la loi) devraient également être incluses.

³À cette fin, le mandat s’appuie sur les définitions figurant dans le document : [Recommandations de l’OCDE sur l’intelligence artificielle](#).

⁴ Voir par exemple : <https://royalsociety.org/topics-policy/projects/data-governance/>

Livrables qui doivent progresser à moyen terme (liste préliminaire discutée en juillet 2020)

Trois examens décrits ci-après, chacun déterminant :

- les exemples internationaux;
- les bonnes pratiques actuelles et les recommandations pour les gouvernements;
- les domaines offrant des occasions de collaboration plus poussées grâce à la recherche de pointe et à l'expérimentation au moyen de projets pilotes;
- les domaines exigeant une enquête plus approfondie de la part du groupe de travail;
- les domaines transversaux de dépendance et de complémentarité entre les trois examens.

(1) Un examen de **l'état des plus récentes approches techniques** concernant la gouvernance des données, abordant, à un niveau élevé :

- la lisibilité des données et des métadonnées par machine, y compris les données sur la provenance et les cartes de vérification des ensembles de données;⁵
- les technologies permettant d'améliorer le respect de la vie privée, notamment les techniques de pseudonymisation et d'anonymisation, l'apprentissage machine fédéré, la confidentialité différentielle et la création de données synthétiques;
- les techniques de détection et de correction des partis pris;
- les technologies qui prennent en charge l'accès aux données, la gestion des contrôles et du consentement (p. ex., les portefeuilles de données individuels), l'enregistrement chronologique des données et la vérification.

(2) Un examen de **l'état des plus récentes approches juridiques** concernant la gouvernance des données, abordant, à un niveau élevé :

- le droit de la propriété intellectuelle tel qu'il s'applique aux données, y compris la collecte/la création de données, la provenance d'ensembles de données, l'utilisation et le partage de données;
- le droit en matière de protection des données, en particulier son application dans le développement de l'AI;
- les mesures légales et réglementaires qui font respecter l'accès aux données et leur réutilisation, notamment l'accès à l'information, la réutilisation des informations émanant du secteur public, l'accès des organismes de statistique, l'accès des villes aux données privées et la portabilité des données;
- le recours à des mécanismes volontaires, à la certification, à la vérification, aux codes de bonnes pratiques, etc. appliqués aux données.

(3) Un examen de **l'état des plus récentes approches organisationnelles/institutionnelles** concernant la gouvernance des données, mettant l'accent sur ce qui suit :

- la souveraineté et la responsabilisation des données individuelles, telles que les magasins de données personnelles, les représentants, les fiducies et les coopératives de données;

⁵ Voir par exemple : <https://arxiv.org/abs/2006.16923>

- la souveraineté et la responsabilisation des données communautaires telles que les fiduciaires de données civiques;
- l'accès aux données pour la recherche, l'innovation et la création de valeur, comme les fiduciaires de données organisationnelles, les clubs;
- la maintenance collaborative des actifs communs;

D'autres résultats sont axés sur des examens plus approfondis et sur la formulation de recommandations qui pourraient viser des résultats particuliers décrits ci-dessus. Par exemple :

- Établir la portée et piloter la création d'ensembles de données spécifiques, représentatifs et ouverts qui soutiennent le développement de systèmes d'IA.
- Procéder à un examen de la documentation, des compétences et des formations requises pour les personnes travaillant dans le domaine de la gouvernance des données.
- Mettre l'accent sur des cas d'utilisation pratiques et propres aux secteurs et décrire un processus de gouvernance des données de bout en bout pour ces cas d'utilisation, en déterminant des exemples, des bonnes pratiques et des recommandations spécifiques pour les gouvernements, y compris des domaines d'harmonisation potentielle.

viii) Annexe 2

Comités directeurs de projet

Le Rôle des données dans le contexte de l'IA

Jeni Tennison (coprésidente), Maja Bogataj Jančič (coprésidente), Takashi Kai, Dewey Murdick, Shameek Kundu, Alejandro Pisanty Baruch, P J Narayanan

Cadre de gouvernance des données

Christiane Wendehorst, Jeni Tennison, Maja Bogataj Jančič, Bertrand Monthubert, Takashi Kai, Shameek Kundu, Alejandro Pisanty Baruch, Te Taka Keegan, Kim McGrail, Josef Drexl