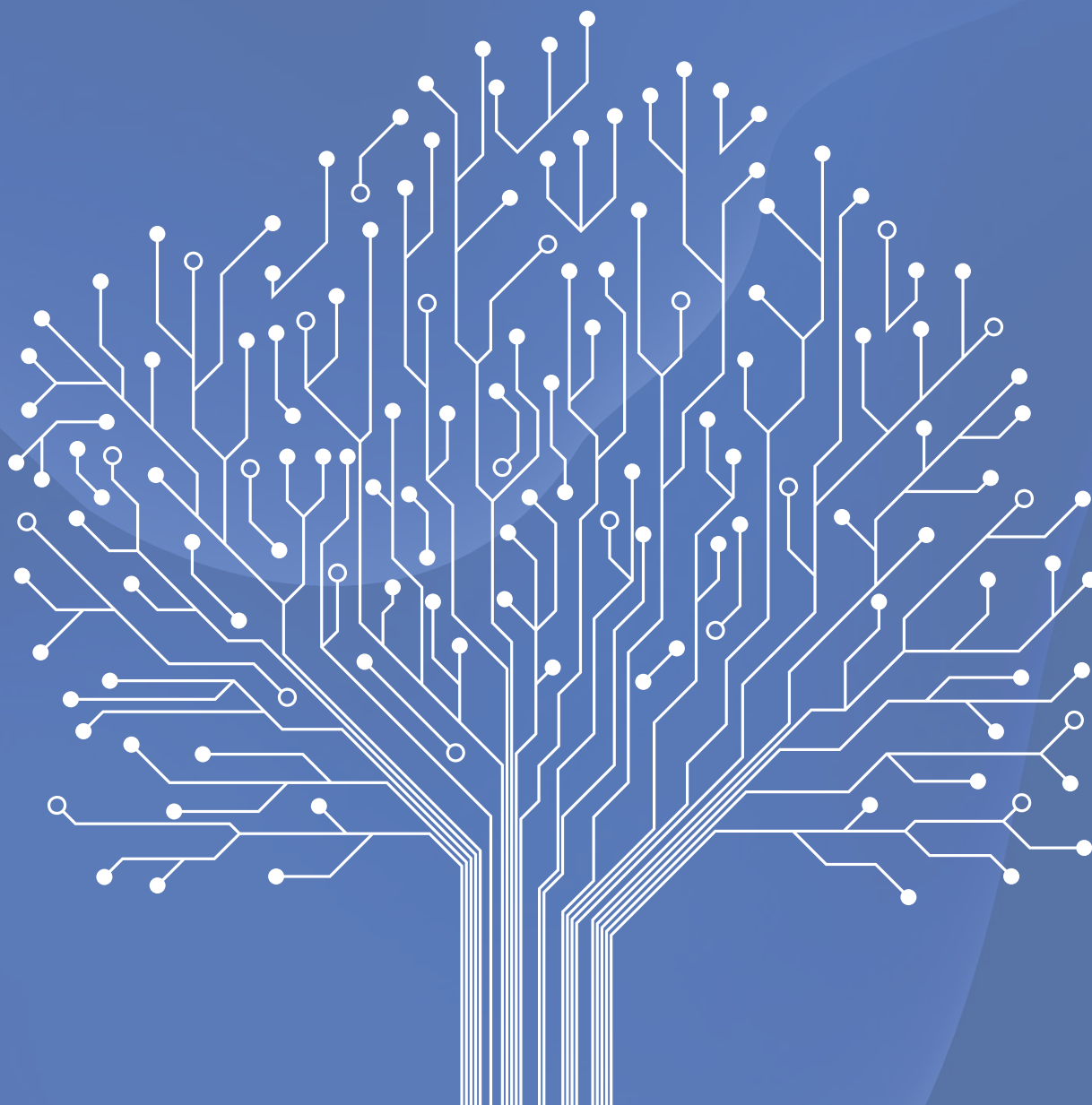


Valoriser les données du secteur public en Écosse et en Europe :

la gouvernance des données à des fins de développement économique, environnemental et social

Rapport du projet



University
of Dundee



scottish universities
insight institute
mobilising knowledge for a better Scotland



University of
Strathclyde
Glasgow

Auteur·rice·s

- 1 Angela Daly**
Université de Dundee
- 2 Esperanza Miyake**
Université de Strathclyde
- 3 Andrew Porter**
Université de Strathclyde
- 4 Anna Berti Suman**
Libre université internationale des études sociales Guido Carli de Rome
- 5 Niamh Nic Daeid**
Université de Dundee
- 6 Stephanie Switzer**
Université de Strathclyde
- 7 Lucille Tetley-Brown**
Université de Dundee
- 8 Julián Valero Torrijos**
Université de Murcia

M^{mes} Daly, Miyake et M. Porter ont dirigé la rédaction de ce rapport avec la contribution des autres co-auteur·rice·s énuméré·e·s.

Remerciements

Ce document est le rapport final du programme du Scottish Universities Insight Institute (SUII) intitulé *Valuing Public Sector Data in Scotland and Europe : Data Governance for Economic, Environmental and Social Development* (« Valoriser les données du secteur public en Écosse et en Europe : la gouvernance des données à des fins de développement économique, environnemental et social ») (Appel à contributions thématiques 2022-2023 : L'Écosse en Europe). Nous remercions le SUII d'avoir financé et soutenu ce travail. Nous remercions également l'ensemble des conférencier·ère·s, président·e·s et délégué·e·s ayant participé à nos ateliers en 2023 et à notre atelier hybride final en janvier 2024 ; nous sommes notamment reconnaissant·e·s à Tim Murray-Browne d'avoir présenté son œuvre lors de notre soirée ouverte au grand public à Glasgow. Nous remercions Glasgow Media Co-op d'avoir mis en page ce rapport et Atlas Translations d'avoir révisé la version anglaise et fourni les traductions en français et en espagnol. Nous remercions Strathclyde Print Services pour les versions imprimées de ce rapport.

Pour citer ce document : Daly, A., Miyake, E., Porter, A., Berti Suman, A., Nic Daeid, N., Switzer, S., Tetley-Brown, L. et Valero Torrijos, J. *Valoriser les données du secteur public en Écosse et en Europe : la gouvernance des données à des fins de développement économique, environnemental et social*, (Série n° 3 Sciences, politique et législation), Université de Dundee, 2024.

Table des matières

Résumé	03
1 Introduction et éléments de contexte	04
2 Ateliers et éclairages	08
Atelier 1 – Les données publiques au service du développement social en Écosse et en Europe, jeudi 5 octobre 2023 (en ligne)	08
Atelier 2 – Les données publiques au service du développement économique en Écosse et en Europe, jeudi 9 novembre 2023 (en ligne)	09
Atelier 3 – Les données publiques au service du développement environnemental en Écosse et en Europe, jeudi 7 décembre 2023 (en ligne)	10
Atelier hybride – La valorisation des données publiques à des fins de développement social, économique et environnemental, 25-26 janvier 2024, Scottish Universities Insight Institute	11
Thèmes transversaux	11
3 Principales recommandations à l'intention des décideurs	14

Résumé

À l'ère du numérique, les données sont présentes en masse autour de nous.

Leur collecte est assurée par un large éventail de personnes et d'organisations, notamment le gouvernement et de grandes entreprises, mais aussi par des personnes ordinaires. Les données offrent de nombreux avantages et perspectives pour améliorer la société, l'économie et l'environnement ; à ce titre, elles peuvent donc répondre aux besoins de publics divers et pluriels, à condition de bien comprendre leur nature et leurs utilisations possibles en vue d'appréhender ce contexte. On peut ainsi commencer par préciser l'éventail des difficultés auxquelles elles donnent lieu, notamment en termes de définition, ainsi que leur grande complexité.

Le programme « Valoriser les données du secteur public en Écosse et en Europe : la gouvernance des données à des fins de développement économique, environnemental et social », financé par le Scottish Universities Insight Institute (aujourd'hui fermé) entre octobre 2023 et janvier 2024, a donné lieu à une étude opportune des avantages et des difficultés associés aux données publiques (du secteur public). Il a permis d'évoquer et d'analyser les nuances que recouvre le fait d'attribuer une valeur aux données, ainsi que les contributions positives – et négatives – que pourraient apporter ces données pour le grand public. Nous avons examiné comment et par qui les « données publiques » sont collectées et utilisées en Écosse, dans l'Union européenne et ailleurs, en identifiant les principales informations susceptibles d'éclairer les travaux de recherche et les processus d'élaboration des politiques à venir dans ce domaine.

Le présent document offre quelques éléments de contexte relatifs à notre programme sur les données publiques, avant de proposer un aperçu des ateliers que nous avons organisés et des principaux enseignements que nous en avons tirés. Dans ce contexte, nous présentons un ensemble de

recommandations à l'intention des décideurs politiques, regroupées selon les différents thèmes apparus pendant notre programme, à savoir :

1 Définitions de la notion de « données publiques »

2 Points de vue du grand public sur l'utilisation des données

3 Valeur des données

4 Maîtrise du numérique

5 Données de santé, données financières et données relatives au système de justice pénale

6 Partage des données

7 Considérations relatives à l'IA



Introduction et éléments de contexte

Entre 2022 et 2023, dans le cadre de son programme intitulé *Unlocking the Value of Data* (UVOD, « Libérer la valeur des données »), le gouvernement écossais s'est associé à un groupe d'expert·e·s indépendant·e·s (GEI) pour animer une discussion consacrée à l'accès du secteur privé aux données personnelles du secteur public en Écosse.

En tant que présidente (M^{me} Daly) et membre (M^{me} Miyake) de ce GEI, nous avons souhaité poursuivre la réflexion sur certaines des questions urgentes que ce travail a permis de soulever, au-delà même de l'Écosse, des données personnelles et de l'accès du secteur privé à ces données, à savoir :

Comment définir les données publiques (du secteur public) ?
Comment sont-elles utilisées ?
Quelle·s valeur·s y est·sont attachée·s ?

Ces questions clés orientent le présent rapport, qui est le principal résultat d'un programme financé par le Scottish Universities Insight Institute (SUII) : *Valuing Public Sector Data in Scotland and Europe: Data Governance for Economic, Environmental and Social Development* (« Valoriser les données du secteur public en Écosse et en Europe : la gouvernance des données à des fins de développement économique, environnemental et social »). Une série d'ateliers organisés en 2023 et 2024 a réuni des personnes issues du monde universitaire, des décideurs politiques et des membres de la société civile, ainsi que des spécialistes et des représentant·e·s du grand public. Dans le cadre de diverses activités, nous avons présenté des études, des perspectives stratégiques et des bonnes pratiques concernant la meilleure façon de gérer les données du secteur public et d'en tirer parti de manière appropriée et éthique dans les contextes écossais et européen.

Notre programme a examiné la question des données du secteur public, de leurs utilisations possibles et de leur gouvernance au service du développement économique, environnemental et social, dans une perspective comparative entre l'Écosse et l'Europe. Ce programme a essentiellement porté sur l'étude des différences entre l'Écosse et l'Europe en matière de pratiques, de politiques, de définitions et de cultures relatives aux données publiques, en particulier dans le contexte post-Brexit ; toutefois, il a également pris en compte les éclairages intersectoriels, transfrontaliers et interdisciplinaires issus des autres nations du Royaume-Uni et du reste du monde, au-delà de l'Europe (notamment de l'Algérie, du Brésil et des États-Unis).

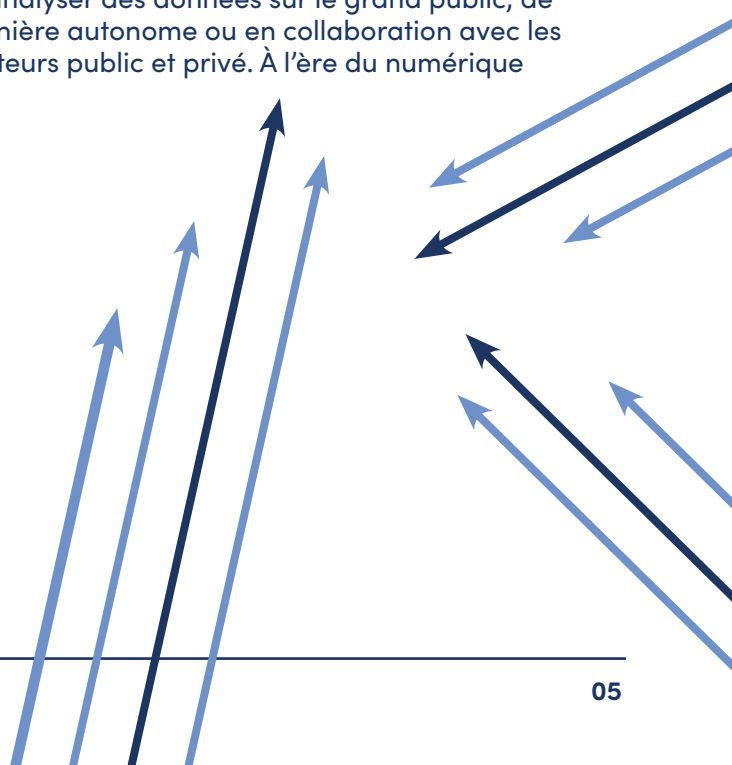
La collecte et l'utilisation des données du secteur public par les gouvernements – à des fins de recherche, d'innovation et de développement, par exemple – sont un enjeu central à l'échelle internationale. En effet, dans un certain nombre de cas de figure administratifs et démocratiques, la richesse et l'exhaustivité des données collectées par le secteur public sont indispensables au renforcement des services publics. Si ces questions sont importantes, c'est aussi en raison de la valeur potentielle des données pour des utilisations ultérieures et secondaires dans les domaines de la recherche, de l'innovation et du développement, que ce soit dans les secteurs public ou privé, dans le troisième secteur ou dans le milieu universitaire. En d'autres termes, la valeur des données publiques réside dans leur capacité de transformation globale des communautés, des économies, des technologies et de notre environnement.

Dans le contexte écossais, le retrait du Royaume-Uni de l'Union européenne (Brexit) a eu des répercussions importantes sur les sphères juridique et politique, notamment en ce qui concerne les services et les données numériques.

D'un groupe à l'autre, la valeur des données publiques donne lieu à des points de vue, à des objectifs et à des motivations différents, voire opposés, d'où la nécessité d'ouvrir la discussion sur cette notion de « valeur ». Ainsi, l'accès des autorités publiques à des données détenues par d'autres parties prenantes, en particulier dans le secteur privé, donne lieu à certaines difficultés. La possibilité, pour des organismes publics, d'accéder à d'autres sources de données (et de s'appuyer sur ces dernières), telles que les sources du troisième secteur ou les initiatives de sciences participatives englobant de nouvelles approches méthodologiques, comme les données participatives (sur la base du respect des règles éthiques et du consentement), est une question qui soulève des difficultés similaires. Comment concilier de telles problématiques ? Ces enjeux trouvent un écho particulier dans l'actualité politique de l'Écosse, avec le programme UVOD du gouvernement écossais mentionné plus haut, ainsi que les récentes évolutions de l'UE, notamment la loi sur les données (Data Act) et la loi sur la gouvernance des données (Data Governance Act), qui créent un cadre pour le développement d'espaces de données sectoriels regroupant des données publiques, privées et générées par les citoyen·ne·s.

La première difficulté porte sur la définition de la notion de données publiques (du secteur public). Le concept de secteur public peut varier d'un pays à l'autre et s'appuyer sur des positions idéologiques quant à la nature des services et activités qui devraient être fournis par l'État et/ou par le secteur privé. Les marchés publics et les partenariats public-privé engendrent d'autres complications, car des « services publics » peuvent être fournis

en tout ou en partie par le secteur privé. De fait, ces subtilités sont prises en compte par la directive de l'UE sur les données à travers la définition des termes suivants : « organisme du secteur public », « organismes de droit public » et « entreprises publiques ». La situation se complique encore lorsque des services privés sont utilisés massivement par le public, tels que les réseaux sociaux propriétaires ou les plateformes en ligne (TikTok, Amazon, etc.), qui collectent, monétisent et surveillent de grandes quantités de données sur les utilisateur·rice·s, tout en offrant peu de transparence ou de précisions quant à leur utilisation. D'autres parties prenantes, telles que les organisations caritatives et les organisations non gouvernementales (ONG), peuvent collecter et analyser des données sur le grand public, de manière autonome ou en collaboration avec les secteurs public et privé. À l'ère du numérique



Des questions se posent, cependant, lorsque ces données génèrent une valeur financière : qu'advient-il de cette dernière ? Le public – ou plutôt, les divers publics – bénéficie-t-il d'une part des gains financiers ?

marquée par la surabondance des données, il est difficile d'établir avec certitude quelles sont les données admissibles ou accessibles et dans quel contexte elles le sont, ainsi que d'identifier les personnes en mesure d'arbitrer et d'en décider – une difficulté susceptible de remettre en question les cadres de référence établis. La propriété intellectuelle et le secret commercial empêchent parfois le public de prendre connaissance des modalités d'utilisation des données. Par ailleurs, les notions habituellement associées à la propriété intellectuelle sont bouleversées par les initiatives de biens communs numériques, qui peuvent également donner lieu à de nouvelles façons d'organiser les données et les biens publics numériques, au-delà du concept habituel d'organisation du secteur public ou privé. Plutôt que de limiter notre analyse aux données du secteur public, nous utilisons le terme plus large de « données publiques » pour couvrir toutes ces possibilités.

Dans le contexte écossais, le retrait du Royaume-Uni de l'Union européenne (Brexit) a eu des répercussions importantes sur les sphères juridique et politique, notamment en ce qui concerne les services et les données numériques. Pour le moment, le Royaume-Uni adhère au règlement général sur la protection des données (RGPD) de l'UE en transposant ces normes dans sa propre législation. La nouvelle législation de l'UE – notamment la loi sur les données et la loi sur la gouvernance des données, qui contribuent à l'instauration de grands espaces de données publiques-privées-citoyennes, et la loi sur l'IA, qui réglementera l'intelligence artificielle – ne fait partie ni du droit ni de la politique du Royaume-Uni. En conséquence, les approches du Royaume-Uni et de l'UE en matière de gouvernance des données divergent. Il convient également de noter que le Royaume-Uni

adopte une approche très proactive à l'égard des données et des avantages attendus, comme on peut le constater dans les récentes politiques adoptées par le gouvernement, avec notamment la stratégie nationale en matière de données. Il reste maintenant à établir si l'arrivée au pouvoir d'un nouveau gouvernement travailliste au Royaume-Uni donnera lieu à des modifications fondamentales de la politique du pays en matière de données.

En outre, au Royaume-Uni, en vertu de la dévolution du pouvoir (forme de décentralisation qui consiste à donner de l'autonomie aux gouvernements de certaines nations composant le Royaume-Uni, comme l'Écosse), certaines questions telles que la santé relèvent de la compétence d'administrations autonomes, dont fait partie le gouvernement écossais. Ce point a été clairement mis en évidence au début de la pandémie de COVID-19 quand des mesures ont été prises pour en freiner la propagation. En effet, ces mesures variaient quelque peu d'une nation à l'autre du Royaume-Uni, ce qui a eu des répercussions numériques : les applications de suivi des cas contacts différaient d'une administration autonome à l'autre et ne collectaient pas les mêmes données. C'est encore le cas actuellement : bien que la pandémie se poursuive, selon l'Organisation mondiale de la Santé (OMS), les données collectées par les gouvernements ont fortement diminué : au Royaume-Uni, seul le gouvernement écossais poursuit une surveillance complète de la population à travers l'analyse des données relatives aux eaux usées. L'Écosse dispose également de plusieurs niveaux locaux de gouvernement, qui fournissent leurs propres services numériques au public et recueillent des données sur ce dernier de plusieurs façons, ce qui complique encore la définition et l'appréhension de la notion de « données publiques ».

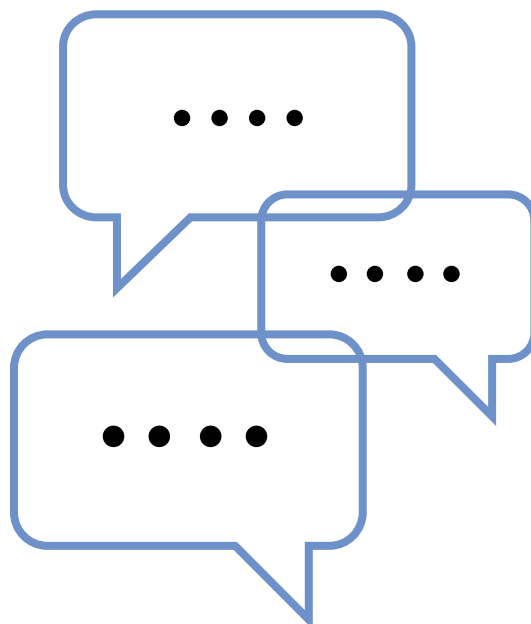
Ces difficultés soulèvent plusieurs questions importantes : pourquoi les données publiques sont-elles collectées ? Quelles données sont collectées ? Quelle est la valeur de cette collecte ? Qui est responsable des décisions prises en la matière ? Les données publiques ont une grande « valeur », à la fois pour de bonnes et de mauvaises raisons : même si elles sont utilisées dans le domaine de la santé (pour de « bonnes » raisons), elles peuvent avoir des effets néfastes lorsqu'elles sont collectées au mépris de l'éthique (pour de « mauvaises » raisons). Bien qu'il soit impossible d'évaluer objectivement et indépendamment de tout système de valeurs ce qui est véritablement et incontestablement « bon » ou « mauvais » au sein de sociétés pluralistes, la signification de cette notion de « valeur » en matière de données publiques peut être analysée par rapport aux normes sociétales les plus courantes. Nous avons choisi d'utiliser les notions de développements économique, environnemental et social comme points de référence pour mesurer la valeur des données publiques. En ce sens, la notion de valeur est particulièrement vaste et ne doit pas se limiter au domaine financier ou économique. Des questions se posent, cependant, lorsque ces données génèrent une valeur financière : qu'advient-il de cette dernière ? Le public – ou plutôt, les divers publics – bénéficie-t-il d'une part des gains financiers ?

En outre, la notion même de « public » s'avère tout aussi complexe. Nous avons utilisé le terme « publics » au pluriel plutôt qu'au singulier pour prendre acte de deux éléments : la diversité des membres du public et des publics, dans la perspective d'une analyse de leurs interactions avec les données publiques en Écosse et dans l'UE ; leurs divergences de points de vue et d'expériences en matière de données publiques. Ce choix est également pertinent pour les questions relatives aux inégalités dans les données. Les groupes ethniques minoritaires, par exemple, sont parfois sous-représentés ou mal représentés dans les données en raison du manque de données au départ ou de l'emploi de méthodes de collecte insatisfaisantes, comme cela s'est produit en Écosse. Il arrive également que des problèmes surviennent lorsque les publics font l'objet d'une collecte de données abusive et inutile, comme nous l'avons observé dans certains domaines de la police prédictive. Il est important de faire participer les publics aux processus de collecte de données et de ne pas envisager les données publiques sous le seul angle du secteur public ou du gouvernement (en s'intéressant aux initiatives de sciences participatives, par exemple, qui peuvent même aboutir à de nouveaux types de gouvernance et d'organisations participatives, comme la coopérative de données de santé MIDATA, en Suisse).

Les données jouent un rôle important et complexe dans un monde de plus en plus en polycrise,

combinant la pandémie en cours, la perte de biodiversité, la pollution, le changement climatique et les guerres. Certains estiment qu'elles faciliteront notre transition verte, bien qu'elles aient également leurs propres coûts environnementaux. Pour d'autres, elles sont une source de nouvelles possibilités économiques au service du changement social, mais s'accompagnent également de nouveaux défis, en raison des infrastructures financières internationales complexes et/ou des contraintes liées à des contextes géopolitiques et sociaux spécifiques. Ces questions démontrent la nature contingente et politique des données publiques, qui peuvent être utilisées à des fins différentes et parfois contradictoires. Il est donc urgent d'examiner ce qui constitue les « bonnes » et les « mauvaises » utilisations des données publiques, de s'intéresser aux acteur·rice·s responsables de ce type de décision, et d'identifier les parties prenantes qui bénéficient de ces utilisations.

À l'ère de l'IA, l'examen de ces questions deviendra un sujet encore plus pressant : les processus décisionnels algorithmiques utilisés dans le secteur public (police, santé, etc.) aboutiront à une plus grande complexité, voire à une transformation des questions éthiques, des aspects juridiques et des conséquences de l'utilisation des données publiques. En outre, avec l'utilisation généralisée et populaire de l'IA générative, l'un des principaux défis auxquels nous serons confrontés ne concerne pas seulement la façon dont nous collectons et utilisons les données publiques ou dont nous y accédons – ou leur valeur, qu'elle soit « bonne » ou « mauvaise » –, mais aussi la nécessité de questionner avec plus de rigueur les modes de production des données publiques et les fins auxquelles elles sont utilisées. Il est crucial pour la recherche, la pratique et l'élaboration des politiques à venir d'examiner le rôle de l'IA dans la production, la collecte, l'analyse et l'utilisation des données publiques.



Ateliers et éclairages

C'est dans ce contexte que nous avons mis en œuvre notre programme, qui comportait trois ateliers en ligne et un atelier final hybride, organisés entre octobre 2023 et janvier 2024.

Dans le cadre de nos ateliers, les participant-e-s ont cherché à comprendre et à traiter les problèmes susmentionnés liés à la valorisation des données publiques en s'appuyant sur diverses études, perspectives stratégiques et bonnes pratiques ayant trait à la gestion et à l'utilisation des données du secteur public, afin de veiller à ce que leur valeur présente un intérêt public du point de vue social, économique et environnemental.

En partenariat avec d'autres universitaires, décisionnaires politiques, spécialistes et membres du grand public issus de l'UE, ces ateliers ont été l'occasion d'explorer plusieurs sujets :

- a) la place de l'Écosse dans les cadres élargis de gouvernance des données en vigueur au sein de l'Europe post-Brexit ;**
- b) la diversité des différents publics et leurs interactions avec les données du secteur public en Écosse et dans l'UE ;**
- c) les inégalités dans les données, pour assurer un développement économique, social et environnemental à la fois éthique et équitable.**

Les trois ateliers en ligne ont traité de la question des données à des fins sociales, économiques et environnementales, respectivement, et l'atelier hybride final a réuni tous ces thèmes. Nous résumons ici certains sujets et points clés qui ont émergé et qui, avec le contexte décrit ci-dessus, ont éclairé les recommandations formulées dans la section suivante.

Atelier 1 :

Les données publiques au service du développement social en Écosse et en Europe

Jeudi 5 octobre 2023 (en ligne)

Cet atelier a débuté par une série de conférences données par des intervenant-e-s invité-e-s sur le thème des biens communs numériques, du lien entre droits humains et données de santé, de la science des données pour le bien commun, de l'IA dans la santé et du lien entre données et inégalités.

Nous avons discuté des défis liés à l'anonymisation des données, en particulier des données de santé, des besoins en matière de diversité des données et des enjeux publics relatifs à la confiance, à la sécurité et à la confidentialité. Dans les salles de réunion, les discussions ont porté sur ce que nous pouvons considérer comme des données publiques et sur la définition du terme « public », sur la façon dont les données publiques peuvent être utilisées au service du développement social, ainsi que sur les possibilités et les difficultés correspondantes. Les possibilités et difficultés liées à l'utilisation des données publiques au service des évolutions sociales ont aussi été évoquées, notamment la question épineuse de la définition du développement à des fins sociales. Les participant-e-s ont pris acte des difficultés et des nuances associées à la définition des données publiques, des dangers que présentent les grands modèles d'IA et de l'érosion du droit à l'oubli, ainsi que des discours publics relatifs à la façon dont l'utilisation des technologies et des données est envisagée.

Les membres du groupe ont été unanimes : de leur avis, lorsqu'on réfléchit à l'utilisation des données publiques, il est fondamentalement plus utile de s'interroger sur ce que nous pouvons ou non contrôler. Au niveau individuel, il apparaît nécessaire d'assurer une meilleure compréhension de la définition des données publiques et, plus important encore, de la manière dont ces données sont utilisées ou partagées. Les discussions ont également porté sur le caractère évolutif de la valeur des données en tant que ressource, sur les conséquences sociales d'un tel point de vue sur les données (envisagées de manière autonome en tant que « bien »), en particulier les données de santé, et sur la façon dont ce point de vue pouvait contribuer aux problèmes auxquels l'humanité et l'environnement sont confrontés.

Atelier 2 : Les données publiques au service du développement économique en Écosse et en Europe

Jeudi 9 novembre 2023 (en ligne)

Cet atelier a également débuté par une série de discussions – sur l'architecture financière des données ouvertes, sur le rôle des données dans le développement économique, sur l'intelligence collective et sur les répercussions des futurs espaces de données de l'UE. L'IA étant actuellement une préoccupation à l'échelle mondiale, certaines de ces discussions ont également évoqué l'utilisation de l'IA dans le cadre des sciences participatives.

Nous avons discuté des avantages et défis associés à l'intégration de nouvelles données aux données existantes, de la législation européenne actuelle et à venir, et de la manière dont le concept d'« altruisme numérique » – qui désigne la façon dont on utilise les technologies numériques et les données correspondantes pour favoriser le développement de grandes valeurs sociétales communes, en transcendant les perspectives économiques traditionnelles – se matérialise dans les politiques et les processus décisionnels. En effet, la nouvelle loi sur la gouvernance des données de l'UE traduit cette notion par le concept d'« altruisme des données », qui suppose que les personnes et les entreprises autorisent l'utilisation de leurs données dans l'intérêt public, sans aucun coût ni rétribution. D'autres discussions étaient axées sur ce que nous pouvons considérer comme des données publiques et sur la définition du terme « public-s », sur la façon dont les données publiques peuvent être utilisées au service du développement économique, et sur les possibilités et difficultés correspondantes. Les discussions ont également porté sur les différences entre le Royaume-Uni et l'UE depuis le Brexit (et à l'échelle internationale) en matière de données



publiques au service du développement économique. Enfin, il a été question d'identifier les parties prenantes qui assument le coût des données publiques et celles qui bénéficient de ce type de données au service du développement économique.

Lors de discussions de groupe plus larges, les participant-e-s ont souligné :

- les défis que présente l'utilisation des données dans des cadres juridiques spécifiques, ainsi que le coût et la valeur associés à la maintenance des données et aux métadonnées ;
- les difficultés à définir la qualité des données en fonction de leurs conditions spécifiques d'utilisation ;
- la question des données envisagées en tant que bien matériel, qui est souvent au cœur de la notion de développement économique ;
- les cadres intrinsèques et les modes de fonctionnement qui résultent de cette réflexion, en particulier lorsque les données ont peu de valeur intrinsèque en dehors de leurs modalités d'interprétation et d'utilisation.

Atelier 3 :

Les données publiques au service du développement environnemental en Écosse et en Europe

Jeudi 7 décembre 2023 (en ligne)

Lors du troisième et dernier atelier en ligne, consacré au développement environnemental, les discussions ont porté sur le droit d'accès aux informations environnementales dans le cadre de la Convention d'Aarhus de la Commission économique des Nations Unies pour l'Europe, de l'initiative Espace de données pour des villes et communautés intelligentes et durables (DS4SSCC) et du contexte politique de l'UE, avec notamment la directive INSPIRE et la directive sur les données ouvertes.

Les discussions se sont concentrées sur la définition des notions de « données publiques » et de « public:s », sur la manière dont les données publiques peuvent être utilisées au service du développement environnemental, et sur les différences entre le Royaume-Uni et l'UE depuis le Brexit (et à l'échelle internationale) en matière de données publiques au service du développement environnemental. Les possibilités et difficultés

associées à l'utilisation des données publiques à des fins de développement environnemental ont également été abordées, en particulier la question de la contextualisation des données d'un site à l'autre, notamment au niveau local, régional et national, ainsi que dans les zones urbaines et rurales.

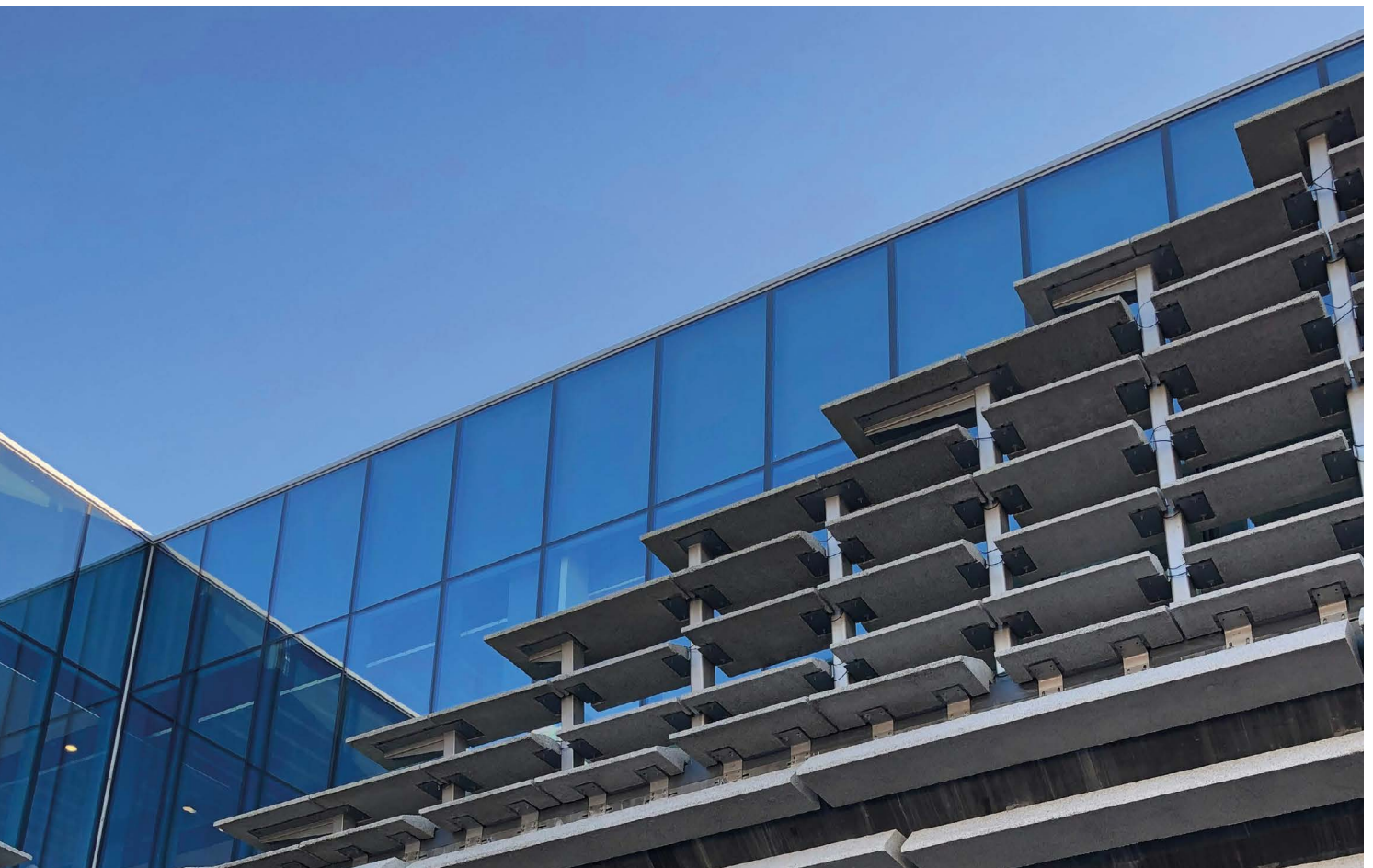
Les participant-e-s ont souligné :

→ **les complexités et les tensions qui apparaissent lors de la création de cadres « universels » dans des zones géographiques présentant des contextes et des besoins différents ;**

→ **la nécessité de donner aux publics et aux organismes publics les moyens de maîtriser le numérique afin qu'ils puissent contextualiser et utiliser correctement les données environnementales ;**

→ **les éventuelles inégalités dans l'accès aux données et les différences d'utilisation des données entre les zones urbaines et rurales à l'échelle locale, régionale et nationale ;**

→ **la nécessité de repenser le partage des données et la manière dont l'information circule, en particulier en ce qui concerne la nature des données que diffusent les organismes publics auprès des différents publics, la nature des données collectées au sein de ces publics et la manière dont les données peuvent être diffusées pour mieux faire comprendre aux publics les besoins et les défis environnementaux.**



Atelier hybride : La valorisation des données publiques à des fins de développement social, économique et environnemental

25-26 janvier 2024

Scottish Universities Insight Institute

Les enseignements tirés des ateliers en ligne consacrés aux données sociales, économiques et environnementales ont été intégrés dans un atelier hybride d'échange de connaissances organisé sur deux jours et réunissant des universitaires et des spécialistes d'Écosse, de l'UE et d'ailleurs.

L'atelier hybride s'est déroulé en personne au Scottish Universities Insight Institute, dans le centre de Glasgow, et en ligne sur la plateforme Zoom, où des participant·e·s du monde entier ont été accueilli·e·s. Durant les deux jours d'atelier, divers points de vue, conférences et activités ont été proposés. Les deux journées ont débuté par un discours liminaire, suivi d'une série de sessions – discussions, table ronde ou encore atelier interactif – avec des universitaires et des spécialistes de différents secteurs et disciplines qui ont fait part de leurs points de vue respectifs et de leur expérience en matière de données publiques.

À la fin de la première journée, un événement public de l'artiste Tim Murray-Browne, intitulé *Joy and Ethics in Making AI Art* (« Joie et éthique dans la production de l'art génératif »), était organisé dans le cadre de l'atelier. Ouvert au grand public, qui a répondu présent, cet événement gratuit a exploré les expérimentations audiovisuelles menées par l'artiste avec des outils d'intelligence artificielle, et son utilisation de ses propres données comme pan intégral de la création de nouvelles œuvres d'art par l'intermédiaire de l'IA. Cet événement a donné une dimension concrète à certaines des questions discutées dans la journée, offrant au public l'occasion de les envisager à travers le travail de M. Murray-Browne.

L'atelier s'est terminé par une table ronde intitulée « De la recherche à la politique », qui rassemblait des membres d'organisations axées sur les données visant à apporter des changements politiques. Les conférencier·ères ont fait part de leur expérience de collaboration avec des chercheur·se·s universitaires et ont formulé des recommandations à l'intention des universitaires dont l'objectif consiste à peser sur l'élaboration des politiques.

Thèmes transversaux

Les trois ateliers ont permis de faire émerger plusieurs thèmes de discussion, qui avaient en commun des préoccupations et des approches similaires relatives au concept de données publiques. Pour certains de ces thèmes, les participant·e·s ont fourni des détails sur les difficultés rencontrées et les solutions envisagées.

Définitions de la notion de « données publiques »

Les participant·e·s ont souligné la subtilité dont il faut faire preuve pour définir les « données publiques » et les défis liés à l'application de telles définitions dans des contextes pluriels. Le fait que ces débats et ces complexités aient été pris en compte et que la question des définitions ait fait l'objet d'un consensus a été jugé productif et utile à l'obtention de résultats complémentaires dans les différents groupes.

En outre, les concepts de propriété et de contrôle sont très difficiles à analyser lorsque l'identité des propriétaires des données n'a pas été établie. Par exemple, dans les organisations du secteur bénévole, on trouve des données dont la propriété n'est pas clairement définie. Ces données peuvent-elles être considérées comme publiques ?

Points de vue du grand public sur l'utilisation des données

Les participant·e·s ont souligné les difficultés associées aux points de vue des différents publics sur l'utilisation des données et ses nuances. On insiste souvent sur le fait que les données sont immatérielles et factuelles, et sur la méconnaissance de l'identité des personnes qui collectent les données et des personnes dont les données sont recueillies. Il est donc nécessaire que le public comprenne mieux qui collecte des données et pourquoi ces données sont collectées.

On part souvent du principe que le partage des données publiques constitue une pratique normalisée. Dans le grand public, certain·e·s hésitent à partager des données en se fondant sur l'hypothèse selon laquelle ces dernières seront partagées avec d'autres personnes. Il s'avère donc nécessaire de mieux faire comprendre aux individus les pratiques de partage de données.

Des exemples ont été puisés dans des stratégies relatives aux données adaptées au contexte local, dans le cadre desquelles les acteurs qui se situent en première ligne s'efforcent d'obtenir les meilleurs résultats. Dans certaines stratégies, les données sont délibérément manipulées pour obtenir les « meilleurs » résultats et ne sont pas représentatives

de la réalité. Cela peut être une source de confusion et de problèmes de compréhension.

Voici quelques exemples clés de problèmes liés aux points de vue du grand public sur l'utilisation des données : les flux de données entre l'État et le public ; les pratiques de partage de données entre les organisations de santé publique et les entreprises privées, qui peuvent susciter la méfiance du public ; dans le cadre de l'économie de la surveillance, les problèmes sociaux au sens large – au-delà des seules questions de vie privée – qui découlent de l'extraction des données des publics par les plateformes numériques.

Maîtrise du numérique

Les participant-e-s aux ateliers ont noté la nécessité d'accroître la capacité des publics à comprendre et à contextualiser correctement les données correspondant à leurs contextes et à leurs besoins respectifs. Ainsi, en raison de paramètres culturels et historiques, les définitions et les utilisations des données peuvent différer d'une zone géographique à l'autre, à l'échelle locale, régionale et nationale. Il est important de se rappeler que les données n'apportent pas de réponses intrinsèques ; à l'inverse, les réponses dépendent du cadrage et des processus d'utilisation des données. Tout repose sur l'utilisation concrète qui en est faite, en particulier s'agissant de l'utilisation étendue des données par les organisations et les gouvernements. Il est également nécessaire de veiller à ce que le secteur public ait une meilleure compréhension des questions relatives à l'utilisation, à la signification et à la contextualisation des données, en gardant à l'esprit qu'en dernier ressort, ce même secteur public est composé d'individus qui n'ont peut-être pas tous le même niveau d'expertise et de compréhension. Entant que fonctionnaires individuel-le-s, les personnes qui travaillent au service des publics doivent avoir les compétences nécessaires pour exploiter au mieux les possibilités offertes par l'ère numérique, tout en restant vigilantes face aux risques de préjudice pour les personnes bénéficiaires de ces services (que ces derniers soient assurés par le gouvernement ou par d'autres entités de service public).

Données sur la santé

Dans l'ensemble des ateliers sur les données sociales, économiques et environnementales, la question des données de santé a fait l'objet d'un intérêt particulier. Les participant-e-s ont notamment relevé le fait que les variations dans l'utilisation et la gestion des données de santé publique pouvaient contribuer aux inégalités en matière de santé. Dans le cadre de l'utilisation des données de santé, il s'avère compliqué de trouver l'équilibre entre l'ouverture des données et le maintien de la confidentialité des données des patient-e-s. L'anonymisation et la protection des données sont essentielles dans certains cas de figure. Les données pseudonymisées présentent aussi des problèmes :

les individus peuvent encore être identifiables si les données sont présentées à grande échelle et avec suffisamment de caractéristiques identifiables.

Il reste d'importants problèmes à résoudre au sujet des données portant sur des caractéristiques susceptibles d'être utilisées pour identifier les individus, telles que des pathologies inhabituelles ou rares, ou des données issues de zones rurales (donc de populations plus petites), car les individus sont plus susceptibles d'être réidentifiés a posteriori. Parmi les autres difficultés rencontrées, on peut notamment citer des exemples de personnes qui ont été identifiées ou rattachées à une identité antérieure susceptible d'avoir changé. Au Royaume-Uni, des efforts considérables sont déployés pour créer des environnements de recherche fiables visant à empêcher la réidentification.

Le fait de regrouper les données de santé à grande échelle (couplage de données) et de créer des systèmes interopérables autour des mêmes données présente des avantages. Des exemples ont été donnés concernant la nécessité d'éviter d'avoir à répéter les mêmes informations lors de consultations médicales. Un traitement plus efficace des données et l'interopérabilité sont essentiels pour aider à surmonter ces obstacles.

Enfin, il a été noté que, dans le domaine de la santé, le respect des processus éthiques et des normes de recherche était pris très au sérieux. Cela souligne la nécessité d'adopter des normes et des règles d'éthique plus globales pour déterminer plus facilement quelles sont les utilisations des données jugées acceptables ou non. La prochaine loi de l'UE sur l'IA donne un exemple complémentaire des différents types de normes et de cadres qui pourraient être mis en œuvre.

Partage des données

Les participant-e-s ont noté que les notions relatives au partage des données et, par conséquent, à la protection des données personnelles, constituaient souvent une source de tension. Les données sont généralement considérées comme des éléments sur lesquels le service public s'appuie pour agir, avant d'être envisagées comme une ressource au service des publics. En particulier, les flux de données entre les publics et l'État ne sont pas totalement équilibrés. Le droit du public à présenter des données est limité à bien des égards, en raison des questions juridiques, de ce que l'État est prêt à accepter, et du manque de ressources suffisantes au niveau de l'État pour traiter les informations et les données communiquées par les publics. La complexité de la situation est renforcée par les attentes en matière de redevabilité et de transparence, qui ne sont pas toujours parfaitement compatibles, en particulier dans les contextes politiques prenant en charge divers publics aux préférences ou aux idéologies parfois opposées. Il est largement admis que les gouvernements et les organismes du secteur public qu'ils administrent sont dominés par une culture de la prudence et

Les participant·e·s ont noté que l'interprétation commune des données en tant que « bien » tangible était souvent au cœur du débat sur le développement économique et considérée comme faisant partie intégrante d'une société de l'information. Cependant, les données sont façonnées et définies par les différentes manières dont elles sont utilisées.

de l'aversion au risque. Cela pourrait ralentir et compliquer la transition vers de nouveaux modes de travail, dans un contexte d'évolution numérique notamment marqué par l'essor des grands volumes de données au cours de la dernière décennie.

L'accent a également été mis à plusieurs reprises sur le concept d'« altruisme numérique », avec pour question centrale la pertinence de cette approche de la politique et de la prise de décision. Dans de nombreux cas, la réciprocité dans le partage et l'utilisation des données est souhaitable. En l'absence de telles exigences, l'échange de données devient un autre mode d'échange. Par exemple, il existe des cas d'utilisation de données privées dans des contextes publics ou dans le cadre de travaux de recherche. Cela crée alors des difficultés relatives à la perception et à la transparence de telles utilisations, ainsi qu'aux implications politiques. L'un des principaux défis consiste à persuader les entreprises privées de se comporter de manière altruiste, en particulier en ce qui concerne l'utilisation, le partage et la gestion des données.

Enfin, il a été souligné que les différentes formes de données ne présentaient pas toutes la même valeur et que les coûts de maintenance des différentes formes de données pouvaient varier. Par exemple, dans certains cas, les métadonnées peuvent avoir une « valeur » plus élevée lorsqu'elles sont utilisées et partagées.

Intégration des données nouvelles et existantes : les participant·e·s ont noté les difficultés et les avantages potentiels associés à l'intégration d'ensembles de données anciennes et nouvelles. Bien que les nouveaux projets puissent souvent révéler de nouvelles sources de données, ils peuvent également mettre au jour de nouvelles façons

d'interpréter et de travailler avec les données existantes. Ces nouvelles méthodes d'interprétation peuvent augmenter la valeur d'utilisation des données existantes et, en conséquence, bénéficier aux personnes et aux communautés auprès desquelles celles-ci ont été obtenues. En outre, les participant·e·s ont noté que les nouvelles données pouvaient être utilisées pour vérifier et améliorer les données existantes. L'efficacité de l'utilisation des données consiste souvent à faire dialoguer les formes de données nouvelles et existantes. Cette démarche est envisagée dans les futurs espaces de données de l'UE, un modèle émergent de partage de données nouvelles et existantes susceptible d'avoir des répercussions majeures sur la façon dont les ensembles de données publics sont combinés et utilisés.

Perceptions des données en termes objectifs

Les participant·e·s ont noté que l'interprétation commune des données en tant que « bien » tangible était souvent au cœur du débat sur le développement économique et considérée comme faisant partie intégrante d'une société de l'information. Cependant, les données sont façonnées et définies par les différentes manières dont elles sont utilisées. La conceptualisation des données en tant que bien concret va de pair avec la nécessité de toujours collecter et créer plus de données.

En outre, il est important de garder à l'esprit la non-fongibilité des données au sein des biens communs numériques, afin de ne pas percevoir ces dernières uniquement comme une propriété privée ou un « bien » traditionnel. Lorsqu'on met l'accent sur leurs caractéristiques objectives, les données ne sont pas toujours de haute qualité au regard des conditions d'utilisation qui leur sont associées. En dernier ressort, seules comptent l'utilisation efficace des données et les raisons qui motivent cette utilisation.

Principales recommandations à l'intention des décideurs

Sur la base de notre programme, nous avons élaboré les recommandations suivantes, en particulier à l'intention des décideurs politiques travaillant sur les données publiques.

1. Définitions de la notion de « données publiques »

Nous recommandons aux décideurs politiques de prendre acte des défis liés à la définition des notions de « données du secteur public » ou de « données publiques ». Dans cette perspective, il convient de s'interroger sur l'identité des parties prenantes incluses dans la notion de « public » (il peut exister plus d'un type de public ou de publics), ainsi que sur d'autres aspects conjoncturels et structurels (des différences régionales ou environnementales peuvent exister). En outre, nous recommandons aux décideurs politiques de mettre ces questions en relation avec les concepts de contrôle et de propriété des données, et de prendre part aux discours publics et à la réflexion sur l'utilisation des données et des technologies.

2. Points de vue du grand public sur l'utilisation des données

Nous recommandons aux décideurs politiques de s'interroger sur le sens que les publics eux-mêmes donnent à la notion de « données publiques » et sur la façon dont ils perçoivent la collecte de leurs données, les acteurs responsables de cette collecte et les raisons qui la motivent. Cette réflexion devrait éclairer la manière dont les décideurs politiques s'efforcent de renforcer la maîtrise et les différentes interprétations, par les publics, de la notion de « données publiques », des acteurs responsables de leur collecte, des raisons qui motivent cette dernière et des modalités d'utilisation et de partage des données.

3. Valeur des données

Nous recommandons aux décideurs politiques d'identifier et d'évaluer la signification de la « valeur » des données publiques, en tenant également compte de la façon dont les différents publics valorisent les données. Une première étape de ce processus devrait consister à identifier les normes et les repères utilisés pour mesurer la « valeur » des données publiques. Nous encourageons vivement à : évaluer cette « valeur » en ne se limitant pas à la valeur financière ou économique ; appréhender ces valeurs dans un contexte plus large, comme la valeur sociale ou environnementale ; accepter la complexité de ces sujets, plutôt que de chercher à simplifier à outrance, au mépris de toute nuance.

4. Maîtrise du numérique

Nous recommandons aux décideurs politiques de chercher à accroître la capacité des publics à comprendre et à contextualiser correctement les données correspondant à leurs contextes et à leurs besoins respectifs. En outre, nous recommandons aux décideurs de veiller à ce que le secteur public ait une meilleure compréhension des questions relatives à l'utilisation, à la signification et à la contextualisation des données, en gardant à l'esprit qu'en dernier ressort, ce même secteur public est composé d'individus qui n'ont peut-être pas tous le même niveau d'expertise et de compréhension.

5. Données de santé, données financières et données relatives au système de justice pénale

Notamment pour les données de santé et la santé, nous recommandons aux décideurs, spécialistes et chercheur·se·s de prendre acte de la difficulté à trouver un équilibre entre l'ouverture des données et le maintien de la confidentialité des données des patient·e·s, ainsi que des défis relatifs aux données portant sur des caractéristiques susceptibles d'être utilisées pour identifier les individus. Dans le même temps, il convient de garantir l'intégration et l'interopérabilité des données, le cas échéant, afin d'éviter la nécessité de répéter les mêmes informations lorsque les individus consultent plusieurs professionnel·le·s de santé. Il va de soi que le fait de devoir comparer la valeur potentielle de l'utilisation de données privées au profit des publics et le préjudice susceptible d'y être associé concerne également d'autres types de données sensibles, telles que les données financières et les données du système de justice pénale. Un tel équilibre demande de grandes compétences. Nous recommandons donc aux décideurs et aux spécialistes d'autres secteurs des données publiques de tirer les enseignements des processus d'éthique et des normes de recherche déjà en place – notamment dans le domaine de la santé – afin d'adopter des normes et des règles d'éthique plus globales, permettant de déterminer plus facilement quelles sont les utilisations des données jugées acceptables ou non.

6. Partage des données

Les décideurs et les spécialistes doivent prendre acte des tensions associées au partage des données, notamment en ce qui concerne les déséquilibres de pouvoir et de ressources entre les publics et l'État, et entre les publics et les entreprises. Les décideurs politiques devraient également envisager des moyens de parvenir à un plus grand « altruisme des données » au sein du secteur privé, afin que les acteurs du secteur privé partagent leurs données avec le secteur public et le grand public.

7. Considérations relatives à l'IA

Nous recommandons aux décideurs, spécialistes et chercheur·se·s de prendre acte de l'utilisation de plus en plus répandue de l'IA dans la production (la création), l'analyse et l'utilisation de données (publiques), et de réfléchir à ses conséquences. Cette démarche devrait permettre de dresser un tableau général des coûts et des avantages associés à l'IA, du rôle que doivent jouer les données publiques dans les écosystèmes d'IA (en tenant compte de ce qui est utile pour les publics) et des questions de propriété et de contrôle de l'IA.



University
of Dundee



scottish universities
insight institute
mobilising knowledge for a better Scotland



University of
Strathclyde
Glasgow