

**Commission économique pour l'Europe****Comité de l'énergie durable****Groupe d'experts des systèmes  
de production moins polluante d'électricité****Dix-neuvième session**

Genève, 3 et 4 octobre 2023

Point 7 de l'ordre du jour provisoire

**Fiabilité et cyber-résilience des systèmes  
énergétiques intégrés intelligents****Groupe d'experts  
de l'efficacité énergétique****Dixième session**

Genève, 5 et 6 octobre 2023

Point 6 de l'ordre du jour provisoire

**Transition numérique et résilience  
des systèmes énergétiques****Améliorer l'efficacité et la fiabilité des systèmes énergétiques  
grâce à l'analyse des mégadonnées****Note du secrétariat***Résumé*

Source de perturbations, l'évolution du secteur de l'énergie vers, entre autres, la décarbonisation, la décentralisation et la transition numérique alimente la transition énergétique et a des répercussions majeures sur le secteur des services publics de distribution dans le monde entier. Compte tenu de ces tendances, l'intelligence artificielle en général et, plus spécifiquement, l'analyse des données s'avèrent donc indispensables.

Les services publics de distribution et les fournisseurs d'énergie recourent de plus en plus à l'analyse des mégadonnées, à l'apprentissage automatique et à l'intelligence artificielle, et ce, à un rythme qui pourrait être supérieur à celui de leur maturation. De fait, ils peuvent avoir déjà atteint un stade avancé de déploiement pour ce qui est des algorithmes. Pourtant, faute de stratégie, ils peuvent ne pas avoir élaboré de plans réalistes pour la conservation des données, les ensembles de données d'entraînement ou les résultats de l'analyse des données.

Le présent document est le résultat d'une collaboration entre le Groupe d'experts des systèmes de production moins polluante d'électricité et le Groupe d'experts de l'efficacité énergétique, menée sur la plateforme de l'Équipe spéciale de la transition numérique dans le domaine de l'énergie. Il s'agit d'une version étoffée du document non officiel portant la cote GEEE-9/2022/INF.3 intitulé « Policy discussion – Challenges of Big Data and analytics-driven demand-side management » (Concertation sur les mesures à prendre – Problèmes posés par les mégadonnées et la gestion de la demande au moyen d'analyses), dans lequel figure la série originale de questions posées à l'Équipe spéciale de la transition numérique dans le domaine de l'énergie, pour un examen plus approfondi en collaboration avec les organes subsidiaires du Comité de l'énergie durable.

La mention d'une entreprise, d'un produit, d'un service ou d'un procédé breveté n'implique aucune approbation ni critique de la part de l'Organisation des Nations Unies. Les appellations employées ne reflètent en aucun cas une quelconque prise de position du Secrétariat de l'Organisation des Nations Unies quant au statut juridique de pays, territoires, villes ou zones quelconques, ou de leurs autorités.



## I. Introduction

1. Parmi les technologies qui sous-tendent la transition numérique du secteur des services publics de distribution figurent des systèmes énergétiques intégrés axés sur la production et la consommation d'énergie distribuée sous forme d'énergie solaire photovoltaïque et éolienne, la production et le stockage d'énergie à l'échelle des services, les véhicules électriques et l'infrastructure de recharge de ces véhicules, ainsi que la multiplication des réseaux électriques intelligents dotés de compteurs intelligents intégrés et d'autres infrastructures et équipements à commande numérique.

2. Toutes ces nouvelles technologies produisent des données. La multiplication de ces technologies numériques s'accompagnera d'une augmentation de la quantité de données à recueillir, à traiter et à analyser. Cette tendance à l'omniprésence des mégadonnées crée de nouveaux débouchés pour les systèmes informatisés d'aide à la décision extensifs et robustes. Et pourtant, faute de ressources et de connaissances spécialisées, les travaux de recherche sur les mégadonnées et l'intelligence artificielle n'en sont qu'à leurs premiers balbutiements dans le secteur des services publics de distribution de l'énergie alors que dans d'autres secteurs, tels que le commerce en ligne et les télécommunications, la recherche sur les mégadonnées et l'intelligence artificielle se développe aussi rapidement que la technologie qui les sous-tend. Pour amorcer l'intégration de l'intelligence artificielle dans le secteur de l'énergie, il faut pouvoir disposer, pour les algorithmes, d'ensembles de données rétrospectives portant au moins sur quelques années. Par conséquent, pour utiliser l'intelligence artificielle globalement, les services publics de distribution doivent commencer à recueillir des données dès les toutes premières phrases du processus d'analyse.

3. Ainsi, ces technologies émergentes permettront de créer de nouveaux modèles d'activité, de nouveaux systèmes énergétiques intégrés intelligents et des capacités au sein des services publics de distribution, de renforcer la participation des consommateurs et de créer de nouveaux comportements, en particulier s'agissant de la demande. Les nouveaux systèmes énergétiques intégrés qui seront établis, associés à un grand nombre de sources d'énergie renouvelable et à la production de variables, ne permettront d'obtenir un équilibre entre l'offre et la demande que s'ils reposent sur les mégadonnées et l'intelligence artificielle. S'il dispose d'un soutien adéquat dans le domaine de la recherche, du financement et du soutien stratégique, le secteur des services publics de distribution peut faire de la collaboration internationale et de la concurrence loyale une réalité au sein de cet espace technologique.

## II. Contexte

4. L'expression « analyse des données » découle des travaux réalisés depuis l'apparition du terme « Business intelligence » (aujourd'hui traduit par « informatique décisionnelle »), et se trouve, depuis cent ans, au cœur de l'évolution des travaux dans le domaine de l'informatique (voir l'annexe). Elle est aujourd'hui utilisée dans presque tous les secteurs industriels et commerciaux. Plusieurs facteurs sont responsables de l'intérêt actuel pour l'analyse des données dans le secteur de l'énergie. Grâce à la baisse du coût des technologies de l'information et de la communication ainsi qu'à l'augmentation de la puissance de calcul, les données sont disponibles en plus grand nombre et de nouvelles possibilités d'analyse ont vu le jour (facteur d'incitation). En outre, le caractère toujours plus fini des sources d'énergie renouvelable et la nature dynamique de l'offre due à l'arrivée constante de nouveaux acteurs sur le marché renforcent la complexité et créent de nouveaux besoins en matière d'analyse des données (facteur d'attraction)<sup>1</sup>.

5. Depuis 2005, le nombre de publications sur l'intelligence artificielle augmente de manière significative dans tous les domaines, y compris l'énergie et l'informatique, mais en 2010, les articles sur l'analyse des données dans le secteur de l'électricité, en tant que domaine de recherche, ont même dépassé ceux sur l'intelligence artificielle.

---

<sup>1</sup> Frederik vom Scheidt *et al.*, « Data analytics in the electricity sector – A quantitative and qualitative literature review », *Energy and AI*, vol. 1 (2020).

6. Le nombre de définitions de l'intelligence artificielle n'a d'égal que celui des applications relatives à son utilisation. Dans le présent document, on entend par « intelligence artificielle » la technologie de pointe qui se fonde sur l'analyse de données pour automatiser le processus de prise de décision concernant les stratégies de participation de la clientèle, exploiter au mieux les prévisions relatives à la consommation d'énergie et aux flux d'énergie pour la production et le stockage à l'échelle locale, améliorer la détection des vols et des fraudes, échanger des produits de base grâce à des prédictions plus précises, et gérer et sécuriser efficacement le réseau énergétique contre les cyberattaques avant qu'elles ne se produisent.

7. L'analyse de mégadonnées est une application de l'intelligence artificielle qui nécessite des compétences en matière de conservation, de gestion et d'analyse des données. Le rôle d'un analyste de mégadonnées va généralement au-delà de celui des analystes traditionnels spécialistes de l'informatique décisionnelle. Dans le présent document, on entend par « analyse des mégadonnées » l'analyse d'un ensemble de données à l'aide d'algorithmes et d'autres techniques avancées de modélisation et d'analyse statistique afin de produire, à partir de ces données, des informations utiles exploitables. L'expression connexe « analyse avancée » est souvent décrite comme étant l'utilisation d'analyses prédictives et prescriptives (parfois également appelées intelligence artificielle) pour transformer ces informations utiles en actions. Dans le présent document, l'analyse avancée porte sur le mesurage et la gestion des événements du réseau et de la demande des clients. Toutefois, il est intéressant de noter que les analyses avancées effectuées dans les entreprises peuvent entraîner une réduction des coûts d'exploitation de 8 % à 9 % (aspects de l'analyse des données relatives aux personnes : amélioration des enquêtes sur les accidents du travail, gestion et prévention, recrutement, formation, gestion des performances, fidélisation des employés, etc.).

8. L'augmentation des capacités des ordinateurs et de la puissance de calcul s'est accompagnée d'une hausse de la quantité de données collectées, stockées et traitées quotidiennement. L'augmentation du nombre de centres de données, grâce à l'avènement de l'Internet (et du Web), signifie que 2,5 quintillions d'octets (2,5 trillions de gigaoctets, Go) de données sont créés tous les jours.

9. Bien que la communauté scientifique ne dispose pas de définition normalisée des mégadonnées, 3 à 10 caractéristiques se dégagent. Le terme « mégadonnées » ne désigne pas seulement le volume de données, mais aussi la vitesse de transmission élevée et la grande variété d'informations qu'il est difficile de collecter, de stocker et de traiter à l'aide des technologies classiques. Dans le présent document, on entend par « mégadonnées » des ensembles de données hétérogènes extrêmement volumineux provenant d'une grande variété de nouvelles sources de données que les logiciels de traitement de données traditionnels ne peuvent pas traiter en temps voulu (en temps quasi-réel).

### III. Problèmes recensés

#### A. Partage et démocratisation des données

10. De plus en plus, être connecté à Internet devient un besoin quotidien, non seulement pour le côté pratique de la connexion, mais aussi pour l'accès aux données et aux informations nécessaires aux entreprises pour atteindre leurs objectifs. En tant que technologie, la connectivité n'est qu'un point de départ qui, dans le pire des cas, peut exacerber les obstacles existants ou en créer de nouveaux.

11. Le partage et la démocratisation des données sont fondamentaux pour le concept d'inclusion numérique, défini comme étant l'utilisation, la gestion et la conception équitables, concrètes et sûres de technologies et de services numériques et l'exploitation des possibilités qui y sont associées, pour tous et partout<sup>2</sup>. Si l'on veut qu'elles soient adoptées et exploitées au mieux, ces nouvelles technologies et les ensembles de données qui leur sont associés doivent être utiles et fiables.

<sup>2</sup> Voir : [https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/general/Definition\\_Digital-Inclusion.pdf](https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/general/Definition_Digital-Inclusion.pdf) (consulté le 7 mai 2023).

12. Pour que les données soient faciles à exploiter, elles doivent être partagées par de nombreuses parties prenantes, y compris celles qui se trouvent en marge d'un secteur ou d'une société, et les problèmes de cybersécurité, de confidentialité, de propriété et de respect de la vie privée doivent être résolus. La traduction en fonction du pays et de la région constitue également un problème (pertinence linguistique). Il convient de noter que l'accès aux formations sur les compétences nécessaires à la transition numérique est étroitement lié à la présence des langues locales sur le Web. Les principaux bénéficiaires de ces compétences transversales sont les services publics de distribution et les fournisseurs d'énergie, que le fournisseur d'énergie soit ou non le producteur ou le propriétaire des infrastructures du réseau.

13. L'accès intégral en temps voulu aux données pertinentes sur la consommation et les clients est un problème qui doit encore être réglé, en particulier dans les domaines dans lesquels la transition numérique est considérée comme le « prochain moteur de croissance ». Les pays d'Europe centrale et orientale illustrent bien la manière dont le passage au numérique (et le partage des données) peut avoir des répercussions à l'échelle mondiale. Ainsi, grâce à l'actualisation, en 2016, du Règlement général européen sur la protection des données (RGPD), des lignes directrices précises sur la manière de traiter les différents types de données ont été établies. Pourtant, sans renforcement de la coopération et de la coordination des politiques entre les différentes régions, il n'est pas possible de concrétiser tous les avantages de la transition numérique et des mégadonnées et d'en tirer parti.

14. Étant entendu que la « démocratisation des données » est un processus continu qui vise à permettre à toutes les parties prenantes, quel que soit leur niveau de connaissances techniques, de traiter rationnellement des données et de prendre des décisions éclairées sur la base de celles-ci, des problèmes devant être analysés de manière systématique perdurent.

### Conservation des données

15. On entend par « conservation des données » le processus de collecte, d'organisation, de caractérisation, de nettoyage, d'amélioration, d'optimisation et de préservation des données. Les données optimisées en vue d'une utilisation analytique doivent être structurées selon leur exploitation et les algorithmes.

16. L'infrastructure de compteurs avancés (AMI), qui compte parmi les plus importantes applications de l'Internet des objets pour les services publics de distribution, offre des avantages aux services publics dans les domaines de l'exploitation et de la clientèle. Même pendant la pandémie de COVID-19, le déploiement des systèmes de compteurs avancés et des compteurs numériques à l'échelle mondiale s'est poursuivi, voire s'est intensifié. Le nombre de compteurs installés devrait dépasser 227 millions d'unités en 2026 dans l'Union européenne (par rapport à 150 millions d'unités en 2020) et les livraisons annuelles de compteurs électriques intelligents en Amérique du Nord passeront de 8,8 millions d'unités en 2019 à 19,9 millions d'unités en 2024<sup>3</sup>. La pénétration des compteurs intelligents en Asie-Pacifique était de 69 % en 2019 et devrait atteindre 82 % en 2025. Les dix marchés qui connaîtront la croissance la plus rapide entre 2020 et 2026 se situeront tous en Europe centrale, en Europe orientale et dans le Sud-Est de l'Europe<sup>4</sup>.

17. Le déploiement de compteurs intelligents n'est toutefois pas limité à l'électricité. Ainsi :

a) Les régions confrontées à un stress hydrique important ont besoin de moyens innovants pour gérer et contrôler l'utilisation de l'eau<sup>5</sup>. Selon les estimations, 700 millions de compteurs d'eau intelligents devraient être connectés d'ici à 2030, par rapport à 196 millions à la fin de l'année 2021. Les zones géographiques où les déploiements devraient être les plus importants d'ici à 2030 sont la Chine (31 % du total), l'Amérique du Nord (29 %) et l'Europe (28 %)<sup>6</sup> ;

---

<sup>3</sup> Nicholas Nhede, « Smart meter penetration in North America will reach 81% by 2024 », Smart Energy International, 5 juillet 2019.

<sup>4</sup> Berg Insight AB, « Smart Metering in Europe – 17th Edition », octobre 2021.

<sup>5</sup> World Resources Institute, « 17 Countries, Home to One-Quarter of The World's Population, Face Extremely High Water Stress », 6 août 2019.

<sup>6</sup> Transforma Insights, « Water Smart Meters: 700 million connections by 2030 to solve issues related to water scarcity and loss », 17 août 2022.

b) Les déploiements de compteurs de gaz intelligents augmentent également à mesure que les technologies de connexion et de connectivité s'améliorent pour les produits de base, ainsi que pour soutenir les politiques gouvernementales visant la construction d'infrastructures pour une distribution et une utilisation efficaces du gaz naturel résidentiel et industriel. Les compteurs de gaz naturel, qui transmettent des relevés différents de ceux des compteurs d'eau et d'électricité, indiquent la pression, le volume et la température du gaz, ce qui permet d'obtenir une autre perspective de l'utilisation dans les locaux connectés. Comme pour les autres produits de base, les compteurs de gaz naturel peuvent donner des informations sur des événements inattendus, lesquelles peuvent être mises en corrélation avec les relevés des compteurs d'autres produits pour donner une vue d'ensemble de l'utilisation et de l'état des locaux (et éventuellement de la santé et de la sécurité des clients)<sup>7</sup> ;

c) Les technologies connectées que sont les synchrophaseurs (PMU) et les systèmes de surveillance et d'acquisition de données (SCADA) couplés à des infrastructures de compteurs avancées sont des technologies habilitantes essentielles pour les réseaux électriques intelligents, qui fournissent de précieuses métadonnées (« données sur les données »). Les nouvelles technologies de connectivité, telles que les communications massives de type machine (mMTC) 5G, les technologies LPWA (réseau de longue portée à basse consommation) non mMTC et la téléphonie cellulaire 4G devraient remplacer les réseaux maillés radiofréquence, qui sont actuellement la principale technologie de communication utilisée.

18. Le déploiement de ces systèmes permet aux services publics de distribution de tirer parti de la puissance du mesurage à distance pour les services de connexion et de déconnexion, la prévention des pannes et la surveillance de la consommation d'énergie, grâce à une perspective, à très fort niveau de granularité, de l'architecture du réseau ainsi que de l'état et du fonctionnement des infrastructures. Les coûts d'installation de ces systèmes sont relativement faibles, mais les coûts d'exploitation sont élevés, en particulier ceux liés au personnel dûment qualifié (expérience) et formé (éducation).

19. Le simple volume de données produit par ces compteurs intelligents (et d'autres dispositifs de l'Internet des objets) crée une large surface d'attaque, qui devient toujours plus difficile à surveiller et à sécuriser. À titre d'exemple, les renseignements permettant d'identifier une personne (PII) sont définis comme étant toute représentation d'une information qui permet de déduire raisonnablement, par des moyens directs ou indirects, l'identité d'une personne à laquelle l'information s'applique<sup>8</sup>. La définition générale des PII fait l'objet d'un consensus dans l'ensemble du secteur, mais il existe différentes manières de les sécuriser, de les traiter et de les utiliser. De nos jours, les clients sont parfaitement conscients de l'importance de la confidentialité des données et de la nécessité d'une protection personnelle puisqu'ils prennent toujours plus à cœur leur rôle dans la transformation du réseau intelligent. Motivées par la nécessité de protéger le public, de nombreuses réglementations relatives à la protection de la vie privée ont été adoptées récemment, notamment le règlement RGPD et la loi californienne sur la protection des consommateurs (CCPA). Ces mesures viennent compléter des dispositions prises de longue date en matière de sécurité des données dans d'autres secteurs tels que la santé, la finance et le commerce.

20. En outre, les données que les organisations collectent, traitent et stockent dans le cadre de leurs activités courantes mais qu'elles n'utilisent pas à d'autres fins, également appelées « données obscures »<sup>9</sup>, représentent un potentiel inexploité pour les entreprises. En appliquant les principes FAIR (faciles à trouver, accessibles, compatibles et réutilisables)<sup>10</sup> à la conservation des données, les organisations peuvent transformer ces « données obscures » en infrastructures précieuses, lesquelles pourraient être utilisées pour améliorer l'efficacité du réseau et la prévision de l'offre et de la demande d'énergie.

<sup>7</sup> Reports and Data, « Smart Gas Meter Market [...] Forecast to 2028 », juillet 2023.

<sup>8</sup> NIST Special Publication 800-79-2, « Guidelines for the Authorization of Personal Identity Verification Card Issuers (PCI) and Derived PIV Credential Issuers (DPCI) », juillet 2015.

<sup>9</sup> Gartner, « Dark Data », Information Technology Glossary. Disponible à l'adresse suivante : <https://www.gartner.com/en/information-technology/glossary/dark-data> (consulté le 27 mai 2023).

<sup>10</sup> GO FAIR Initiative. Disponible à l'adresse suivante : <https://www.go-fair.org/fair-principles/> (consulté le 29 mai 2023).

21. Dans le contexte de l'analyse de la demande des consommateurs d'énergie, de la segmentation de la clientèle, du regroupement des sources d'énergie et d'autres travaux de profilage de la demande, les modèles d'analyse doivent pouvoir s'appuyer sur le plus grand nombre de renseignements possible pour interpréter la matrice de confusion. Il convient néanmoins encore de déterminer comment l'analyse de la clientèle peut être réellement exploitable lorsque l'information est limitée en raison de contraintes réglementaires concernant des données potentiellement sensibles. À mesure que le secteur de l'énergie progresse sur la voie de la transition numérique, il convient de trouver un équilibre entre, d'une part, l'accès aux données nécessaires pour déterminer et différencier les besoins des différents groupes de population en matière de demande d'énergie et, d'autre part, la confidentialité des données requise pour éviter que des informations sensibles ne soient rendues publiques.

### **Disponibilité des données**

22. Dans le secteur des services publics de distribution, on entend par « disponibilité des données » le processus par lequel les entreprises de services publics (à la fois les opérateurs de systèmes de distribution et les fournisseurs, dans les zones géographiques où les services sont regroupés) et leurs utilisateurs disposent de données continues, sûres et facilement utilisables, associées à leur consommation d'électricité ou d'énergie. Alors que le secteur de l'énergie n'a pas fait preuve d'immobilisme pour ce qui est de l'adoption des technologies numériques, ainsi que de la collecte et de la gestion des données produites, la disponibilité des données est relativement faible. En d'autres termes, bien que les nouvelles centrales électriques construites selon les principes de la transition numérique garantissent une plus grande efficacité et une meilleure disponibilité des services, et grâce à l'essor des jumeaux numériques qui peuvent être utilisés pour la modélisation, la prévision et les tests visant une performance optimale, l'accès aux données interdisciplinaires et intersectorielles fait encore gravement défaut.

23. Si les mégadonnées ne sont pas facilement disponibles et si elles ne sont pas accompagnées de données relatives au réseau électrique, les possibilités d'analyse avancée des données sont considérablement réduites et celle-ci est souvent impossible à mettre en œuvre. Il existe toute une série de fournisseurs de systèmes de gestion des données de mesurage, ainsi que de technologies de données d'utilisation orientées client. Ces systèmes sont souvent incompatibles en raison de logiciels et de formats de données propriétaires.

24. L'objectif des *FAIR Guiding principles for scientific data management and stewardship* (principes FAIR de gestion et d'intendance des données scientifiques) est de fournir des lignes directrices permettant d'améliorer la capacité d'action des machines sur les données, afin d'accroître la réutilisation et l'utilisabilité des données et de faire face aux attributs toujours plus importants que sont le volume, la vitesse et la variabilité des données.

### **Intégration des données et gestion intelligente de l'énergie**

25. Lorsque des données sont créées, collectées et conservées, il y a production d'un historique ou d'une provenance. La provenance donne des indications sur l'origine des données, les modalités de création de l'ensemble de données, les opérations appliquées, les entités qui les ont effectuées, ainsi que la date et la justification de ces opérations.

26. Comme les systèmes et technologies de gestion des données sont déployés de manière hétérogène, selon un jalonnement aval et avec des architectures logicielles différentes, l'intégration des données entre les systèmes énergétiques est souvent limitée, ce qui, dans la plupart des cas, est dû à des problèmes de provenance.

27. En outre, les entreprises de services publics utilisent encore du matériel et des logiciels dépassés, qui sont souvent à la limite de l'obsolescence. Les anciennes technologies se retrouvent donc isolées des nouvelles plateformes, car elles sont incapables d'interagir avec la nouvelle génération de matériel et de logiciels. Le secteur des services publics de distribution, en particulier, est confronté à ces problèmes, car les investissements portant sur la transition numérique sont principalement axés sur les nouveaux systèmes et les nouvelles technologies de réseau (poteaux, câbles, transformateurs, unités de gestion du réseau, etc.), plutôt que sur la mise à niveau des anciens systèmes de gestion des données.

28. D'un point de vue plus technologique, les plateformes de données intégrées doivent étayer la prise de décision dans un environnement hybride, soit un environnement dans lequel il y a intégration avec les applications internes à l'organisation et qui est susceptible de collaborer avec les environnements des nuages publics. Le cycle de vie des données est complexe et varié, du point de collecte jusqu'au consommateur final. Pour ces aspects de l'hybridation et de la gouvernance, il convient de disposer d'une solution de gestion des données réfléchie, correctement mise en œuvre et uniformisée.

29. Les systèmes intelligents de gestion de l'énergie associent des dispositifs d'utilisation finale, des ressources énergétiques distribuées et des systèmes avancés de contrôle et de communication à des modèles normalisés de gestion des données. La mise en place d'architectures de données en vue d'optimiser l'intégration et l'analyse des systèmes est un défi permanent. Il faut encore définir des approches novatrices en matière d'analyse des mégadonnées et considérer l'intelligence artificielle en tant que processus organique au sein des systèmes intelligents de gestion de l'énergie, si l'on veut assurer la pérennité des systèmes et une intégration solide des données.

30. S'agissant des moteurs potentiels de l'intégration des données, pour le secteur des services publics de distribution, ces cas d'utilisation peuvent aller de la modélisation des probabilités de panne à l'amélioration de l'expérience des clients. Bien que les différents cas d'utilisation opérationnels et orientés clients présentent des défis uniques, le fait de pouvoir s'appuyer sur une architecture de données normalisée et sur des processus de gestion des données optimisés pour l'analyse peut accélérer les percées dans le domaine des solutions de pointe<sup>11</sup>.

31. Plusieurs modèles de données normalisés sont disponibles à titre de référence :

a) Aux États-Unis, l'Electric Power Research Institute (EPRI) a publié un cadre synthétique centré sur les services publics de distribution pour l'évaluation de la gestion des données des grandes organisations. L'objectif est de permettre aux cadres supérieurs d'évaluer de manière globale, dans leur organisation, les personnes, les processus et les technologies qui étayaient la gestion et l'analyse des données<sup>12</sup> ;

b) Le modèle commun d'informations statistiques (CIM), développé et géré par DMTF (anciennement Distributed Management Task Force)<sup>13</sup>, est un schéma normalisé international qui fournit une méthode commune pour représenter les éléments de calcul et de réseau d'un système et leurs relations avec d'autres systèmes et éléments, soit un cadre large et robuste pour la gestion des données et des communications d'équipement. Le modèle d'informations statistiques, dont l'utilisation nécessite un ensemble de compétences transversales et une grande expérience, permet de définir et d'organiser une sémantique commune et cohérente pour les équipements et les services, au moyen d'abstractions de classes orientées objet, d'héritages et d'associations de connexions. La gestion de services tels que le diagnostic des pannes, la configuration du système, la comptabilité, la performance et la sécurité est assurée par le modèle commun d'informations statistiques. L'accès aux documents normalisés, aux bibliothèques binaires pour les définitions des classes d'objets et aux hiérarchies de relations est possible après adhésion.

## **B. Compétences disponibles en matière d'analyse au sein des services publics de distribution**

32. Dans le contexte d'un réseau de distribution, les données sont générées et collectées à partir de compteurs AMI, de compteurs intelligents, de stations météorologiques, de systèmes SCADA et de systèmes de transport (recharge de véhicules électriques, etc.). Pour chacun de ces types de système, il existe une multitude de fournisseurs, qui ont chacun leur propre méthode de stockage et de partage des données.

<sup>11</sup> Brad Gall, Chad Tucker, Beth Massey, « Shared Services Common Data Model to Deliver Advanced Analytics », *Proceedings of 2022 IEEE International Smart Cities Conference* (Pafos, Chypre), 2022, p. 1 à 5.

<sup>12</sup> Electric Power Research Institute, « Data Governance and Utility Analytics Best Practices », 30 avril 2014.

<sup>13</sup> DMTF, « Common Information Model », <https://www.dmtf.org/standards/cim> (consulté le 23 mai 2023).

33. Les données haute résolution fournies par des technologies telles que les compteurs AMI sont déterminées par les repères sélectionnés dans le compteur. On peut estimer qu'il est important de collecter autant de données que possible, mais il est plus utile de collecter les données d'intervalle contenant le plus d'informations intégrées. Par exemple, les compteurs intelligents transmettent des relevés de tension qui peuvent également fournir des informations sur l'état et la charge du transformateur associé. Selon le fournisseur, le nombre de repères choisis dans le compteur peut augmenter la quantité de données collectées et les coûts de stockage qui en découlent. Il est justifié d'effectuer, dès le début du processus, une analyse coût-avantages permettant de déterminer la quantité et les types de données les plus utiles. Une telle analyse devrait également tenir compte de l'absence de normalisation entre les fournisseurs, des compétences nécessaires pour se repérer dans les systèmes, variés et hétérogènes, à partir desquels les données sont collectées et des environnements requis pour formater et gérer de manière optimale les données nécessaires à l'analyse, et devrait prendre en compte le travail supplémentaire requis pour l'installation et la mise en service de l'environnement.

34. De nombreux services publics de distribution ne disposent pas de leur propre département d'analyse, et il leur est de plus en plus nécessaire de collaborer entre eux pour déterminer si les algorithmes permettent d'aboutir aux bonnes conclusions. Un fossé sépare le personnel chargé des opérations des services publics, qui connaît le fonctionnement au quotidien, et les analystes de données, qui connaissent les algorithmes. Tous les segments des services publics gagneraient à collaborer afin de limiter la redondance des algorithmes et des ensembles de données.

35. Il est noté que remarquablement peu de choses sont définies en ce qui concerne l'utilisation judicieuse de la technologie. Ainsi, pour les produits qui peuvent être contrôlés par leur fabricant – comme les appareils domestiques intelligents et les nouvelles cartes – les responsabilités du fabricant ne sont pas clairement définies<sup>14</sup>. Les connaissances spécialisées et les politiques d'appui relatives aux préoccupations concernant le respect de la vie privée et la responsabilité des consommateurs en cas de défaillance de l'appareil, qu'elle soit malveillante ou non et fatale ou non, font actuellement défaut, tant au niveau de la conception que de la mise en œuvre.

36. Il convient de mener des recherches supplémentaires dans d'autres domaines clefs, tels que la traduction des données en besoins opérationnels, la monétisation des données et la cybersécurité, si l'on veut régler les problèmes liés à l'amélioration des compétences de la main-d'œuvre actuelle afin qu'elle puisse utiliser efficacement les outils et les techniques disponibles.

### **Traduction des données en besoins opérationnels utile**

37. La traduction des données peut être définie comme étant le processus de conversion de volumes de données d'une syntaxe à une autre et l'exécution de recherches de valeurs ou de substitutions à partir des données au cours du processus. La traduction peut également inclure la validation des données. Par exemple, il peut s'agir de convertir des données de séries temporelles spécifiques à un fournisseur ou même des données de systèmes d'information géographique (SIG) et des fichiers plats de clients, tout en effectuant une validation des données sur les données sources. Pour traduire des données en cibles opérationnelles, il faut disposer d'une vision stratégique à partir de laquelle des objectifs peuvent être établis, car les objectifs mesurables peuvent générer des informations utiles pouvant être exploitées.

38. Dans le secteur des services publics de distribution, les organisations qui travaillent dans le domaine de la production d'énergie traditionnelle (c'est-à-dire à base de combustibles fossiles) doivent ajuster leurs marges de coût en améliorant l'efficacité des centrales. Selon certaines études, une bonne utilisation de l'analyse avancée peut permettre de réaliser des économies de l'ordre de 5 % à 7,5 %<sup>15</sup>. Ces économies peuvent être dues à une amélioration du temps exploitable, à une maintenance prédictive fondée sur la prévision du taux de défaillance ou à une rationalisation de la consommation de carburant associée à un contrôle ciblé des performances visant à réduire (voire éliminer) la surproduction.

---

<sup>14</sup> Jonathan L. Zittrain, *The Future of the Internet – And How to Stop It* (Yale University Press & Penguin UK, 2008).

<sup>15</sup> McKinsey & Company, « The Digital Utility: New challenges, capabilities and opportunities », juin 2018.



39. Selon les estimations de 2018, les technologies fondées sur les données peuvent permettre de réaliser des économies de plus de 12 % sur les coûts d'exploitation et de maintenance. Le coût des capteurs et des dispositifs de capture des données a considérablement diminué, certains ne valant qu'un dixième du prix d'il y a dix ans. À mesure que les progrès en matière de communication (5G, futures technologies 6G, etc.) continuent d'augmenter les vitesses de transmission (1 000 Gbps pour la 6G par rapport à 600 Mbps pour la 5G), les données des solutions de l'Internet des objets peuvent être analysées et une aide à la prise de décision peut être apportée en temps réel.

40. La nature dynamique de ce niveau d'analyse des données entraîne une évolution de la définition des cas d'utilisation stratégique. En tête de liste des services publics qui ont développé leurs programmes de capture de données à haute résolution figurent les cas d'utilisation suivants : surveillance de la probabilité de défaillance, détection et prévision des pannes, sécurité du réseau intelligent, détection des vols, gestion des systèmes d'énergie transactionnelle, maintenance préventive des équipements, optimisation de la performance des infrastructures, gestion de la réaction à la demande, tarifs, facturation en temps réel et amélioration de l'expérience client.

41. Il reste à déterminer si un service public de distribution peut tirer parti des données collectées pour formuler une stratégie articulée autour des besoins connus (ou inconnus).

### **Monétisation des données**

42. Dans une large mesure, la monétisation des données est l'objectif idéal pour parvenir à un déploiement à grande échelle de l'analyse des données. Sans le savoir, les services publics de distribution et les services énergétiques possèdent une mine d'informations très précieuses sur les clients et les opérations, et de plus en plus de données sont disponibles grâce au déploiement d'appareils intelligents.

43. La monétisation des données est l'étape de la maturité des données au cours de laquelle les services publics de distribution et les autres fournisseurs d'énergie exploitent les mégadonnées pour créer de nouvelles possibilités de revenus. Par exemple, tirer parti des informations utiles exploitables à partir des données des utilisateurs et des comportements des clients peut amener un service public à améliorer ses relations avec les clients et à repenser l'expérience de ces derniers. L'utilisation de méthodes telles que les profils de clients à 360 degrés permet de faire face à l'augmentation des taux de défection, qui atteignent 25 % sur certains marchés<sup>16</sup>. Des solutions telles que l'analyse vocale automatisée dans les centres d'appel, intégrée aux systèmes de communication (applications mobiles sur le terrain, etc.) et aux sites Web des entreprises, ainsi que l'analyse de la consommation et la tarification dynamique, permettront aux entreprises d'aller à la rencontre de leurs clients là où ils se trouvent. Ce niveau d'intégration globale augmente la valeur du cycle de vie du produit pour les clients et réduit le taux de défection pour l'entreprise de services publics de distribution.

44. En particulier, pour les services publics qui ne fournissent que des services de distribution, le fait d'améliorer l'efficacité de leur portefeuille de vente grâce à une meilleure évaluation de la solvabilité des clients et des variations de la consommation peut contribuer à réduire au maximum les défauts de paiement et à éviter les fraudes. Selon des études, les stratégies intégrées d'analyse de la clientèle sont susceptibles d'accroître les marges bénéficiaires des entreprises de 5 % à 10 %, tout en augmentant la satisfaction des clients.

45. En outre, de nombreuses données collectées dans le cadre des activités régulières des entreprises ne sont pas utilisées à d'autres fins ; ces données obscures pourraient être importantes pour d'autres entreprises et constituer une nouvelle source de revenus. En s'appuyant sur les informations utiles fondées sur leurs expériences, les services publics peuvent proposer de nouveaux produits et services générateurs de revenus et améliorer les performances des produits et des opérations, afin d'établir des relations plus motivantes et durables avec les clients.

<sup>16</sup> Pablo Boixeda, « Optimizing the Energy Sector with Data Analytics », Cloudera, 20 décembre 2022.

### **Cybersécurité et intégrité du réseau**

46. Alors que le secteur des services publics de distribution privilégie progressivement le numérique, des risques croissants liés à la cybersécurité apparaissent, tant sur le plan opérationnel que commercial. C'est pourquoi, compte tenu des risques élevés de cyberguerre, les services publics de distribution doivent mettre en place des stratégies appropriées de prévention et d'atténuation, tout en élaborant des plans de continuité des opérations en cas de défaillance du dispositif de cybersécurité.

47. Selon un rapport, d'ici à la fin de l'année 2023, les lois modernes relatives à la protection de la vie privée couvriront les informations personnelles de 75 % de la population mondiale<sup>17</sup>. Toujours plus conscients de leur rôle dans l'évolution du réseau dématérialisé, les clients d'aujourd'hui voudront savoir quels types de données sont collectés et quelles en sont les utilisations prévues (et réelles). Les services publics de distribution et les fournisseurs d'énergie, en particulier ceux qui couvrent plusieurs limites géopolitiques dans différentes juridictions, devront disposer d'un programme stratégique d'éducation des clients axé sur la cybersécurité et sur la manière dont les applications sont adaptées à la législation locale.

48. Dans cette même analyse, les auteurs concluent que d'ici 2024, les organisations qui adoptent une architecture maillée de cybersécurité réduiront les répercussions financières des atteintes à la sécurité de 90 % en moyenne. Ce type d'architecture peut être élargi de façon à couvrir les entités hors du périmètre de sécurité traditionnel. L'augmentation du télétravail dans les entreprises va favoriser l'adoption de ces architectures dans les années à venir.

49. Il est également indiqué dans ce rapport que d'ici 2025, un pourcentage toujours plus important d'organisations adoptera des stratégies concernant : les politiques de risque relatives à la cybersécurité pour les transactions avec des tiers (60 %), la législation régissant les négociations sur les logiciels rançonneurs (30 %, par rapport à 1 % en 2021), la nomination d'un représentant du comité de cybersécurité au sein du conseil d'administration (40 %), et la création d'une culture de résilience organisationnelle permettant de survivre aux risques simultanés liés à la cybercriminalité, aux intempéries, aux troubles civils et à l'instabilité politique (70 %).

50. Une analyse plus approfondie et des informations supplémentaires sur les répercussions de la cybersécurité sur le paysage numérique du secteur de l'énergie sont présentées dans le document intitulé « Principales considérations relatives à la cyber-résilience des systèmes énergétiques intégrés intelligents et solutions pour la garantir » (ECE/ENERGY/GE.6/2023/3-ECE/ENERGY/GE.5/2023/3).

### **C. Modélisation de l'analyse des mégadonnées : recherche et développement**

51. L'accès aux travaux en cours dans les domaines de la recherche et de l'application de l'analyse des données à l'échelle des services publics de distribution est un élément essentiel d'intégration de l'analyse des mégadonnées et, de fait, du passage au numérique dans le secteur de l'énergie dans son ensemble. Parallèlement, il est souvent difficile d'accéder aux résultats des travaux de recherche des organisations nationales compétentes et de les exploiter.

52. La maturité des données est un concept important pour les fournisseurs d'énergie. Les services publics de distribution constatent que, de plus en plus, les données sont au cœur de toutes les stratégies d'entreprise, qu'elles contribuent à stimuler l'innovation et qu'elles continuent d'être intégrées dans tous les services ; il est donc nécessaire d'élaborer une stratégie pour l'élaboration d'une feuille de route sur la maturité des données et de la mettre en œuvre. De cette manière, les services publics peuvent rationaliser leurs processus internes, innover et fournir des services fiables à leurs clients.

---

<sup>17</sup> Gartner, « The Top 8 Cybersecurity Predictions for 2021-2022 », 20 octobre 2021.

53. L'évolution de la maturité des capacités en matière de données est représentée dans le tableau ci-après. En passant aux niveaux supérieurs de cette évolution, l'organisation augmente ses capacités de gestion des données, de développement d'algorithmes et de livraison de résultats. Selon des études, les entreprises qui disposent d'une gestion mature des données et de processus analytiques solides (c'est-à-dire reproductibles et vérifiables) peuvent accroître leur rentabilité d'une moyenne de 12,5 % de la marge brute totale<sup>18</sup>. Le passage à un niveau supérieur de cette évolution témoigne de la maturité de la stratégie et de la mise en œuvre de l'analyse des données au sein d'une organisation. Plus les investissements dans les capacités d'analyse sont importants, plus l'organisation peut progresser rapidement. Des études montrent que les investissements, même s'ils sont conservateurs, donnent de bons résultats et que la plupart des entreprises de services publics de distribution et des fournisseurs d'énergie passent d'une phase à l'autre assez rapidement, même lorsque les investissements dans l'analyse des données sont modestes<sup>19</sup>.

Tableau  
**Maturité des données et de l'analyse**

Niveau	Maturité des données et de l'analyse	Capacité d'analyse	Observation
1	Beaucoup de données et trop d'entrepôts de données	Rapports de réaction – questions d'actualité suivies de près <ul style="list-style-type: none"> <li>Listes et imprimés des transactions</li> </ul>	Une organisation a investi dans une infrastructure d'analyse ; mais, il se peut qu'elle soit obsolète ou qu'il n'y ait pas eu de stratégie de gestion des données au départ. Il en résulte que les données sont gérées de manière ponctuelle et que la prise de décision est souvent corrective, en fonction de la priorité du jour et non de preuves empiriques fondées sur la valeur.
2	Rapports de base et automatisation minimale	<ul style="list-style-type: none"> <li>Accent sur les données rétrospectives et le suivi des coûts</li> <li>Pas d'intégration des données ou des applications opérationnelles</li> <li>Les données sont dispersées sur des plateformes de stockage hétérogènes</li> </ul>	
3	Informatique décisionnelle avec analyse statistique	Analyse planifiée – planification à court terme <ul style="list-style-type: none"> <li>Tests diagnostiques</li> <li>Le stockage et l'accès aux données sont automatisés</li> <li>Intégration inter-opérationnelle</li> <li>Compétences analytiques inégales</li> </ul>	Une maturation croissante au sein des départements et de l'ensemble de l'organisation, à mesure que la planification opérationnelle devient plus stratégique et que les compétences en matière d'analyse augmentent.
4	Modélisation prédictive/prescriptive	Analyse stratégique – prestation cohérente <ul style="list-style-type: none"> <li>Fiches et tableaux de bord à l'échelle des départements</li> <li>Capacité d'analyse en cours de maturation</li> <li>Production régulière et sans effort</li> <li>Passage des informations utiles à l'action</li> </ul>	

<sup>18</sup> Richard Carufel, « What's your data really worth? It depends on your data maturity level », Agility PR Solutions, 26 mars 2020.

<sup>19</sup> Electricity Advisory Committee, « Big Data Analytics: Recommendations for the U.S. Department of Energy », février 2021.

Niveau	Maturité des données et de l'analyse	Capacité d'analyse	Observation
5	Optimisation des modèles et des processus	Optimisation des processus – prospective <ul style="list-style-type: none"> <li>• Actions fondées sur la planification future</li> <li>• Intégration et utilisation complètes de données externes</li> <li>• Analyse en temps réel comme facteur de différenciation</li> <li>• Fiches et tableaux de bord à l'échelle de l'organisation</li> <li>• Capacité d'analyse généralisée</li> </ul>	Les processus de l'ensemble de l'organisation sont optimisés et intégrés. Les décisions sont fondées sur la planification future et sur des informations utiles approfondies provenant de prévisions empiriques et solides. Les analyses en temps réel fondées sur des capacités avancées sont déployées de manière cohérente et globale, et les tableaux de bord à l'échelle de l'organisation ont été normalisés, par exemple grâce à des nomenclatures orientées données bien définies.

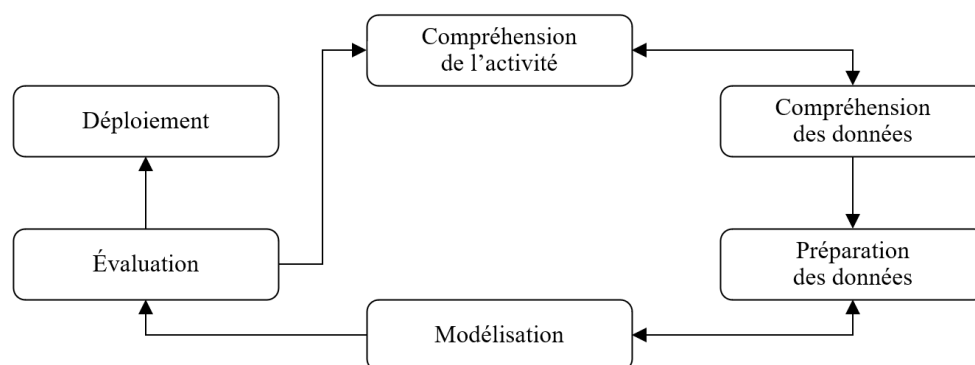
54. Pour passer aux niveaux supérieurs de maturité, les organisations doivent modifier leur paradigme de prise de décision, ce qui nécessite une réflexion intégrée, des applications interfonctionnelles et une collaboration entre les équipes opérationnelles.

#### **Mégadonnées, modèles d'analyse avancée (recherche, développement et déploiement) et sensibilisation**

55. Grâce à la baisse du coût des technologies de l'information et de la communication ainsi qu'à l'augmentation de la puissance de calcul, les données sont disponibles en plus grand nombre et de nouvelles possibilités d'analyse ont vu le jour. Par disponibilité des données, on entend l'accès fiable et en temps voulu aux données pertinentes et l'utilisation fiable et en temps voulu de ces données. Parallèlement, le nombre de sources d'énergie renouvelable et d'autres ressources énergétiques décentralisées continue d'augmenter l'utilisation des réseaux à l'échelle mondiale, ce qui accroît la complexité du système électrique et crée de nouveaux besoins en matière d'analyse de données et de modèles d'analyse optimisés. La disponibilité des données et des modèles d'analyse de données est généralement limitée dans le temps.

56. Les services publics de distribution doivent s'assurer que leurs modèles d'analyse respectent les règles de validité, de reproductibilité et de transparence. Les normes du secteur peuvent représenter une méthode organisée de planification et de mise en œuvre des initiatives d'exploration de données, comme la méthode CRISP-DM (acronyme anglais de « Cross-Industry Standard Process for Data Mining », processus intersectoriel pour l'exploration des données), décrite dans la figure ci-après. Cette méthode est articulée autour de plusieurs phases : compréhension de l'activité, compréhension des données, préparation des données, modélisation, évaluation et déploiement. Les services publics de distribution devraient également envisager d'associer les techniques d'apprentissage automatique et d'intelligence artificielle pour traiter les grands volumes de données complexes sur des plateformes gérées dans le nuage, à des fins d'extensibilité.

Figure  
Cadre de la méthode CRISP-DM



Source : Adapté de <https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview>.

57. Dans un monde où les ingénieurs techniques hautement qualifiés sont toujours plus demandés, il est essentiel que les entreprises accordent la priorité à la formation et au développement de leurs équipes en place. La formation continue et le développement des compétences, l'accent mis sur une culture de prise de décision fondée sur les données, le recrutement d'un mélange d'experts techniques (science des données, apprentissage automatique, statistiques) et d'experts sectoriels (systèmes énergétiques, réglementation), et la promotion de la collaboration et du partage des connaissances entre les membres de l'équipe sont des pratiques clés pour constituer une équipe interne d'analyse qui soit efficace.

58. Les modèles susceptibles de profiter à d'autres, tels que ceux qui anticipent la consommation d'énergie, prévoient la charge et la production des sources d'énergie renouvelable ou analysent la stabilité du réseau, devraient être partagés soit en libre accès, soit pour générer des revenus. En outre, de nombreuses techniques, telles que la régression, l'analyse des séries chronologiques, l'apprentissage automatique (y compris les réseaux de neurones) et les modèles de simulation, pourraient servir de base à ces modèles. Mais, élément le plus important, les ensembles de données et les logiciels libres doivent être accessibles au public, car ils peuvent promouvoir la normalisation et la transparence entre les différentes parties. Le partage de modèles, les techniques de science des données, les données et les logiciels libres peuvent favoriser, de manière non négligeable, une utilisation et une collaboration accrues.

59. Alors que la recherche, le développement et le déploiement de l'analyse des mégadonnées dans le secteur de l'énergie gagnent en ampleur et en diversité, il est essentiel d'intégrer et de structurer le corpus fragmenté des travaux scientifiques. À l'heure actuelle, les analyses de données couvrent les domaines de l'ensemble de la chaîne de valeur, de la production et du commerce à la transmission, à la distribution et à la consommation. Les activités couvrent également différentes applications, telles que la prévision ou le regroupement à l'aide de diverses méthodes telles que les réseaux de neurones artificiels et la création de centres d'innovation régionaux axés sur des technologies données.

60. Des mesures sont prises à l'échelle mondiale pour créer des communautés qui partagent les meilleures pratiques en matière d'utilisation de modèles analytiques avancés. Ainsi, le programme pour une Europe numérique, mis en place par la Commission européenne, souligne la nécessité de créer des projets à fort impact tirant parti de l'intelligence artificielle et de l'analyse de données. Dans le même ordre d'idées, le Ministère de l'énergie des États-Unis d'Amérique a lancé une initiative de modernisation du réseau, afin de mettre en place des réformes révolutionnaires reposant sur l'analyse de données. Afin de partager des informations et de promouvoir l'innovation, les gouvernements forment de plus en plus de collaborations transfrontalières à l'échelle régionale.

61. La collecte de données granulaires complètes, variées et de qualité à partir de nombreuses sources, telles que les dispositifs de l'Internet des objets, les compteurs intelligents et les stations météorologiques, est cruciale si l'on veut améliorer la précision des modèles d'analyse de données. De plus, l'incorporation de flux de données en temps réel, l'utilisation de stratégies d'ingénierie des caractéristiques et l'application d'algorithmes complexes d'apprentissage automatique sont autant d'éléments susceptibles d'améliorer les performances des modèles.

62. La sélection des modèles d'analyse de la demande implique souvent un compromis entre la granularité et la confidentialité, la vitesse et la précision, ainsi que la complexité et l'interprétabilité. La précision, l'interprétabilité et la réactivité des résultats sont directement affectées par ces compromis. Par exemple, un modèle très complexe peut produire des résultats plus précis, mais il peut être plus difficile à comprendre et prendre plus de temps à faire tourner qu'un modèle plus simple.

63. Il convient de tenir compte de l'existence de biais cachés, qui peuvent donner lieu à des données faussées, des hypothèses erronées ou une conception biaisée des algorithmes. Ces biais peuvent entraîner un traitement injuste, un comportement partial ou des informations utiles fausses. Il est donc essentiel de disposer de procédures efficaces de gouvernance des données, qui soient accompagnées de procédures rigoureuses de collecte, de nettoyage, de suivi et de validation des données.

64. Si certains types de biais peuvent être atténués grâce à l'automatisation, il est important de bien comprendre que les algorithmes eux-mêmes peuvent contenir et renforcer des biais, en particulier s'ils sont développés à partir de données biaisées ou déployés sans suffisamment de contrôle. Il est donc essentiel de recourir à des méthodes transparentes et justifiables lors de l'automatisation de tout processus de prise de décision. Les biais algorithmiques peuvent être atténués en utilisant des méthodes justifiables et en respectant des principes éthiques pour ce qui est de l'intelligence artificielle.

#### **Exemple d'analyse avancée et d'application de l'intelligence artificielle pour le repérage des défaillances des compteurs de chaleur**

65. Bien que le potentiel de synergie entre les mégadonnées et l'intelligence artificielle n'ait pas encore été pleinement exploité, des programmes pilotes montrent comment les meilleures pratiques en matière d'analyse avancée peuvent être appliquées, selon des modalités susceptibles d'être adoptées par d'autres.

66. Dans la ville de Vilnius, en Lituanie, le fournisseur de chaleur souhaitait mettre en place un programme d'analyse de la consommation de chaleur des clients au fil du temps afin de repérer les relevés anormaux pouvant indiquer un dysfonctionnement du compteur de chaleur. Statistiquement, 0,5 % des compteurs de chaleur sont catalogués comme étant « en panne » et sont susceptibles de transmettre des relevés d'intervalles erronés. Pour un bassin de 100 000 clients, cela représente 500 compteurs qui envoient des relevés erronés, ce qui peut entraîner des calculs erronés de la demande réelle et, donc, une facturation inexacte.

67. Grâce à des compteurs de chaleur intelligents, des données d'intervalle ont été collectées pendant deux saisons de chauffage. Compte tenu du comportement des clients (basé sur la demande passée) et de la saisonnalité, une analyse statistique des mégadonnées a été réalisée pour éliminer les données non fiables et supprimer les facteurs externes tels que les valeurs aberrantes correspondant aux jours de chauffage hors saison et aux températures non saisonnières, et pour recenser les paramètres d'un scénario de maintien des conditions actuelles. Cet ensemble de données d'entraînement a été utilisé pour entraîner un système fondé sur l'intelligence artificielle, puis pour la facturation des services publics de distribution ainsi que pour repérer les défaillances des compteurs d'énergie thermique.

68. Le système fondé sur l'intelligence artificielle a été testé avec des données mensuelles relatives à la demande de chaleur ; ces données ont été nettoyées et mises à jour pour tenir compte du nombre de jours de chauffage et de la température extérieure pour le mois en cours. Une carte interactive de la consommation réelle d'énergie a été créée à partir des profils des consommateurs de chaleur, puis normalisée en fonction de facteurs d'influence tels que la durée de la saison, la température et la taille de l'appartement. De cette manière, le système fondé sur l'intelligence artificielle peut comparer toutes les données de n'importe quel client, quelle que soit l'année.

69. Les données ont été présentées sur une carte SIG avec une gamme de couleurs allant du vert au rouge et une classe de performance énergétique réelle allant de 1 à 10. Cette carte a été mise à la disposition des clients sur le site Web de la compagnie et a permis de comparer des bâtiments similaires, attirant ainsi l'attention sur les activités d'exploitation, de maintenance et de gestion des bâtiments, dans le but de définir les problèmes potentiels et de les résoudre. Cette façon de procéder est également utile pour l'évaluation et la vérification des mesures d'économie d'énergie, telles que l'isolation des bâtiments et leur rénovation, car les clients peuvent comparer la consommation d'énergie réelle après la rénovation des bâtiments.

70. Bien que cette méthode se distingue des classes de performance énergétique obligatoires A à E, qui n'indiquent que les données de consommation théoriques, elle montre que, sans les analyses de mégadonnées et l'intelligence artificielle, les services publics de distribution auront du mal à recenser correctement les problèmes de consommation d'énergie, à fournir à leurs clients des informations sur la consommation d'énergie réelle et à prendre des décisions à plus long terme en ce qui concerne les dépenses d'investissement et d'exploitation relatives aux infrastructures du réseau.

## IV. Conclusions et recommandations

71. Sur la base des paragraphes qui précèdent, les conclusions et les recommandations suivantes sont présentées pour examen :

- a) Partage et démocratisation des données :
  - i) Conservation des données :
    - Des normes nationales et internationales clairement définies et faciles à comprendre doivent être établies pour le traitement des données sensibles ;
    - Des normes et des opérations prêtes à l'emploi relatives à la cybersécurité et à la sécurité des systèmes cyber-physiques doivent être établies à l'intention des propriétaires de données résidentielles ;
  - ii) Disponibilité des données : l'accès aux données interdisciplinaires et intersectorielles reste très limité. Il convient de poursuivre les travaux de recherche sur l'application des principes FAIR (faciles à trouver, accessibles, compatibles et réutilisables) aux données, notamment les « données obscures » ;
  - iii) Intégration des données et gestion intelligente de l'énergie :
    - Les systèmes doivent être intégrés afin de permettre l'intégration rapide des données provenant de systèmes hétérogènes ;
    - Les normes relatives à la gestion des données doivent être uniformisées ;
    - Des normes et/ou protocoles doivent être établis pour les architectures sous-tendant la gestion intelligente de l'énergie ;
    - Des cas d'utilisation stratégiques doivent être élaborés ;
    - Des architectures de données faciles à comprendre et à utiliser, orientées vers l'analyse et les services de distribution doivent être élaborées ;
- b) Compétences disponibles en matière d'analyse au sein des services publics de distribution :
  - i) Traduction des données en besoins opérationnels :
    - L'analytique avancée s'avère nécessaire en raison de la capture de données haute résolution liée à la multiplication des capteurs et des dispositifs intelligents ;

- La main-d'œuvre existante et les nouvelles recrues doivent être dotées de nouvelles compétences, afin de normaliser les ensembles de données dans les différents services en exploitant au mieux l'augmentation de la puissance de calcul et la robustesse des architectures de données ;
- Il convient de mettre en place des bancs d'essai à grande échelle pour évaluer différentes solutions, puisque le nombre de dispositifs intelligents et connectés de l'Internet des objets continue d'augmenter, ce qui ajoute aux préoccupations des services publics de distribution en matière de formation et de compétences ;
- Il convient de disposer, au stade tant de la conception que de la mise en œuvre, de connaissances spécialisées et de politiques ciblées dans les domaines liés au respect de la vie privée et à la responsabilité à l'égard des consommateurs ;
- Des travaux de recherche plus approfondis doivent être réalisés sur les moyens de perfectionner les compétences de la main-d'œuvre actuelle afin d'utiliser de manière plus efficace les outils et les techniques disponibles et de stimuler les améliorations dans le domaine de la formation ;

ii) Monétisation des données :

- Tarification dynamique pour une augmentation de la valeur économique du client ;
- Modèle potentiel de création de valeur pour les consommateurs en échange de données ;

iii) Cybersécurité et résilience du réseau : les services publics de distribution devront mettre en place un programme stratégique de formation des clients axé sur la cybersécurité et sur l'adéquation des applications avec les réglementations locales ;

c) Analyse et maturité des mégadonnées :

i) Mégadonnées, travaux relatifs aux modèles d'analytique avancée (recherche, développement et déploiement) et sensibilisation :

- Il convient de créer des pôles d'innovation locaux et régionaux afin d'assurer la mise à l'essai et la sécurisation intégrales des technologies numériques et des techniques de traitement des données ;
- Il convient de former les parties prenantes des communautés énergétiques afin d'aider les consommateurs à comprendre et à assumer leur rôle dans la transition numérique ;
- Les parties prenantes du secteur de l'énergie, en particulier les services publics de distribution et les fournisseurs d'énergie, doivent bénéficier d'un plus grand nombre d'incitations concrètes à exploiter les mégadonnées ;
- Il doit être démontré que les cas d'utilisation qui justifient des investissements importants dans la collecte de données, la gestion, l'analyse et l'infrastructure de l'intelligence artificielle garantissent des retours sur investissement bien définis ;
- Le modèle de recouvrement des coûts doit être envisagé de manière à ce que les investissements dans l'analyse des données qui sont justifiables puissent être considérés comme des investissements en capital associés à des avantages clairement définis pour le client.

72. En outre, l'Équipe spéciale de la transition numérique dans le domaine de l'énergie propose les activités complémentaires suivantes :

- a) Étudier les conclusions susmentionnées et mener des travaux exhaustifs et une analyse plus approfondie pour chacune d'entre elles, de préférence en collaboration avec les organes subsidiaires du Comité de l'énergie durable, lesquels pourraient en conséquence élargir le mandat de l'Équipe spéciale de la transition numérique dans le domaine de l'énergie ;



b) Mener des recherches ciblées sur les modèles de financement dans les domaines à privilégier : amélioration des technologies de traitement des mégadonnées (traitement du langage naturel, modélisation des jumeaux numériques, prévisions de la demande et de la charge, optimisation de l'apprentissage automatique, développement des capacités de l'intelligence artificielle, etc.), résilience des réseaux et investissement dans les infrastructures (concernant notamment l'accès aux données et le stockage, la gestion et l'analyse en temps réel des données), conformément au mandat pour 2024 et 2025 de l'Équipe spéciale de la transition numérique dans le domaine de l'énergie.

## Annexe

Tableau  
**Évolution de l'analyse des données**

Année	Événement marquant	Description
1865	Apparition de l'expression « Business intelligence »	Richard Miller Devens utilise l'expression « business intelligence » (renseignement économique, qui se traduit aujourd'hui en français par « informatique décisionnelle ») dans son <i>Cyclopædia of commercial and business anecdotes</i> . Il s'agirait de la première étude sur une entreprise utilisant l'analyse de données à des fins commerciales.
1928	Stockage magnétique	Fritz Pfleumer, ingénieur germano-autrichien, invente un dispositif permettant de stocker des informations sur une bande magnétique. Ses méthodes sont encore utilisées aujourd'hui, la grande majorité des données numériques étant stockées magnétiquement sur les disques durs des ordinateurs.
1928	Ordinateurs de traitement des données	En 1956, IBM lance les 305 et 650 RAMAC (Random Access Memory Accounting), « ordinateurs de traitement de données » dans lesquels se trouvait le tout premier dispositif de stockage sur disque.
1965	Premier centre de données d'un gouvernement	Le Gouvernement des États-Unis d'Amérique prévoit de construire le premier centre de données au monde, dans lequel seront stockés 742 millions de dossiers fiscaux et 175 millions de jeux d'empreintes digitales sur bande magnétique.
1989	Apparition de l'expression « Big Data » (mégadonnées)	L'expression « Big Data » (mégadonnées) est utilisée pour la première fois dans un article publié dans une revue par Erik Larson, auteur de livres de fiction, qui observe que les publicitaires exploitent des données pour cibler les clients.
1996	Rentabilité du stockage numérique	Selon l'ouvrage de R. J. T. Morris et B. J. Truskowski publié en 2003, <i>The Evolution of Storage Systems</i> , le stockage numérique devient moins coûteux que le stockage sur papier.
1998	Next Wave of Infostress (prochaine vague de stress dû à un surcharge d'informations)	John R. Mashey, scientifique principal à SGI, fait une présentation intitulée « Big Data... and the Next Wave of Infostress » (les mégadonnées et la prochaine vague de stress dû à un surcharge d'informations) à l'occasion d'une conférence USENIX.
2001	Apparition de l'expression « 3 V »	Doug Laney définit les trois « V » des mégadonnées : volume, vitesse et variété.
2005	Web 2.0	2005 marque les débuts de Hadoop, plateforme de mégadonnées à code source ouvert, actuellement développée par Apache. Le Web généré par les utilisateurs, connu sous le nom de Web 2.0, est mis au point la même année.
2008	14,7 exaoctets de données fraîches	Dans le monde entier, les serveurs du monde traitent 9,57 zénaoctets (9,57 mille milliards de gigaoctets) de données tous les jours, ce qui correspond à 12 gigaoctets de données par personne et par jour. Cette année-là, on estime que 14,7 exaoctets de données sont créés.

<i>Année</i>	<i>Événement marquant</i>	<i>Description</i>
2009	Directeur de l'informatique (titre orienté données)	Selon Gartner, la principale priorité des directeurs de l'informatique doit être l'informatique décisionnelle. Alors que les entreprises souffrent de l'instabilité économique et de l'incertitude résultant de la Grande Récession, il devient essentiel d'extraire de la valeur des données.
2011	Pénurie des compétences analytiques	Il manque entre 140 000 et 1 190 000 professionnels dotés de capacités analytiques avancées, ainsi que 1,5 million d'analystes et de gestionnaires capables d'émettre des jugements fondés sur des données.
2012	Recherche-développement sur les mégadonnées	Le gouvernement Obama lance la Big Data Research and Development Initiative. Cette initiative axée sur la recherche et le développement dans le domaine des mégadonnées vise à améliorer la capacité à extraire des informations utiles à partir des données, et à accélérer le rythme de développement des sciences, de la technologie, de l'ingénierie et des mathématiques (STEM).
2014	Next Wave of Infostress (prochaine vague de stress dû à une surcharge d'informations)	Aux États-Unis, pour la première fois, le nombre de dispositifs mobiles dépasse celui des ordinateurs personnels de bureau. Deux ans plus tard, en 2016, cette constatation s'applique au reste du monde.
2020	Calcul informatisé en périphérie de réseau (Edge Computing)	Dans le domaine des mégadonnées, la prochaine technique de pointe est le calcul informatisé en périphérie de réseau (Edge Computing), c'est-à-dire le calcul effectué à proximité de la source de collecte des données, plutôt que dans le nuage ou dans un centre de données centralisé.