



MINISTÈRE DE L'ENVIRONNEMENT,  
DE L'ÉNERGIE ET DE LA MER

## Rapport de préfiguration de la fonction de superviseur général des données du ministère de l'environnement

Remis à

Ségolène Royal, Ministre de l'Environnement, de l'Energie et de la Mer, en  
charge des relations internationales sur le climat

Par

Laurence Monnoyer-Smith, Commissaire générale au développement  
durable

Le 31 mai 2016



La France est l'un des premiers États à se doter d'une réflexion, d'un cadre juridique et d'une pratique dans le nouveau champ de possibles ouvert par l'explosion des données. Dans le prolongement des deux lois de 1978, « informatique et libertés » et « accès aux documents administratifs », la loi « informations du secteur public » du 28 décembre 2015 et le projet de loi pour une République numérique stabilisent les concepts et établissent des règles collectives équilibrées.

La fonction d'administrateur général des données, créée en septembre 2014, a commencé à faire la démonstration du bénéfice, pour l'État, d'une pratique renouvelée d'utilisation de ses propres données. Le ministère de l'environnement, de l'énergie et de la mer (MEEM) occupe une place particulière au sein des administrations publiques dans la production de données publiques. Il doit donc assumer les responsabilités que sa position lui confère. La promesse de création de valeur autour de la donnée nécessite un accompagnement résolu dans ce sens. C'est l'objet de la proposition d'instituer une fonction de superviseur des données au MEEM.

# Synthèse du rapport

## § 1. Le ministère de l'environnement est un acteur critique pour la réussite de la politique de la donnée portée par le gouvernement français.

§ 1.1. La France est l'un des premiers États à se doter d'une réflexion, d'un cadre juridique et d'une pratique dans le nouveau champ de possibles ouvert par l'explosion des données. Dans le prolongement des lois cnil et cada de 1978, la [loi](#) « informations du secteur public » du 28 décembre 2015 et le [projet de loi](#) pour une République numérique stabilisent les concepts et établissent des règles collectives équilibrées. La fonction d'[administrateur général des données](#), créée en septembre 2014, a commencé à faire la démonstration du bénéfice, pour l'État, d'une pratique renouvelée d'utilisation de ses propres données.

§ 1.2. Le ministère de l'environnement est un acteur critique pour la réussite de ce mouvement. D'une part, les politiques publiques dont il a la charge — logement, énergie, transport, biodiversité, lutte contre le changement climatique... — voient leurs contextes et leurs écosystèmes profondément modifiés par le développement des pratiques et usages liés au numérique.

D'autre part, parce que les bases de données qu'il gère, directement ou par l'intermédiaire de ses opérateurs — Météo France, ign, onema, ademe, brgm... — sont massives, diverses et centrales pour la vie du pays, voire de la planète (climat ou géodésie, par exemple). Leur production et leur exploitation correspondent à des modèles hétérogènes et toutes n'affichent pas le même degré de robustesse. Dans ce contexte, la Ministre a confié au Conseil général de l'environnement et du développement durable (cgedd) une mission de recensement et de cartographie de l'ensemble de ces données.

§1.3. La création de la fonction de superviseur des données du ministère s'inscrit dans une politique plus globale du ministère qui vise à mettre le numérique au service de la transition écologique. Il s'agit de favoriser l'émergence de nouveaux modèles économiques, de nouveaux produits et services, de valoriser les données du ministère, et de promouvoir une culture partagée autour des enjeux que suscite le numérique dans l'ensemble des aires d'intervention du ministère. La vertu transformatrice du numérique est ainsi considérée comme une opportunité pour accélérer la transition énergétique et écologique et appuyer des évolutions organisationnelles et comportementales nécessaires à son accomplissement.

L'initiative de la *Green Tech verte*, lancée par Ségolène Royal en février 2016, constitue un axe central de cette démarche et l'outil d'opérationnalisation de la politique d'ouverture des données portée par le gouvernement.

## § 2. La richesse des bases de données gérées par le ministère de l'environnement est un potentiel sous-exploité.

§ 2.1. La richesse des bases de données réside dans la multiplicité des usages qui peuvent en être faits. Un premier enjeu essentiel consiste à exploiter le potentiel d'usage par le grand public, rendu possible par l'ouverture et l'accessibilité des données. Qu'il s'agisse de permettre des débats publics informés, de favoriser la production partagée de connaissances ou de susciter la création de services citoyens qui améliorent le fonctionnement de notre démocratie (comme *democratieouverte.com*, on parle alors de *civic tech*), l'ouverture des données contribue à la restauration de la *confiance* et à l'*encapacitation* des citoyens. L'ouverture, prévue par la loi biodiversité, des données des études d'impact est ainsi très attendue. L'implication forte de la France dans le partenariat international pour un gouvernement ouvert (ogp) justifie un effort particulier dans ce domaine.

§ 2.2. Le second enjeu est celui de la création de *nouveaux services* au public permise par l'ouverture de bases de données. Les données de gestion de la voirie d'une ville peuvent par exemple permettre de créer une application de choix d'itinéraires pour personnes à mobilité réduite. Les données d'inclinaison des toitures couplées aux données météo peuvent permettre d'anticiper le rendement de panneaux solaires pour les particuliers. La valeur de ces services peut être marchande ou pas, certains modèles d'affaires restent encore à inventer.

§ 2.3. Que chacun puisse accéder à la donnée dont il a besoin pour exercer sa mission permet également des *économies* et des *gains de productivité*. Au-delà des bénéfices d'un meilleur accès des agents du ministère à la donnée utile, il est aussi possible de mieux allouer les moyens de l'État en s'appuyant sur de nouvelles analyses de données. Ainsi, des modèles prédictifs fondés sur les bases de données décrivant les installations classées et le résultat des contrôles passés peuvent permettre aux agents de l'État de cibler leurs contrôles sur les sites les plus susceptibles d'être en infraction. L'étude fine des consommations électriques des bâtiments administratifs peut permettre de renégocier les contrats d'énergie.

§ 2.4. Enfin, un enjeu essentiel pour l'État concerne son *positionnement stratégique*. La révolution numérique offre un avantage décisif à qui parvient à concentrer l'essentiel des échanges de données entre entreprises et/ou particuliers. Le monopole de fait que détenait l'État sur certains échanges de données (géographie, transports, logements...) est aujourd'hui contesté par de grandes entreprises privées (Google, Uber, Airbnb...). La mise en œuvre d'une politique de la donnée par l'État est donc porteuse de l'affirmation forte de son rôle de régulateur, sous le nom d'État-plateforme, stratégie à laquelle participe pleinement le ministère de l'environnement.

### **§ 3. Plusieurs freins ralentissent la valorisation des données gérées par le ministère de l'environnement et ses opérateurs.**

§ 3.1. Le premier frein tient à la distance qui sépare ceux qui produisent les données (pour un usage déterminé) de ceux qui pourraient les utiliser à d'autres fins. La méconnaissance mutuelle entre producteurs et utilisateurs, même au sein du ministère, illustre le besoin d'une *culture partagée de la donnée*. Se préoccuper de l'usage des données produites par l'exercice quotidien de ses fonctions, ou réciproquement, chercher les données existantes qui permettraient d'augmenter l'impact de son action, doivent devenir des réflexes professionnels au ministère.

§ 3.2. Les *réalités techniques* peuvent également freiner la circulation des données. L'architecture d'un système d'information peut rendre impossible l'extraction de ses données, par exemple. La qualité des données peut être adaptée pour les usages courants, mais s'avérer insuffisante pour d'autres usages à plus forte valeur-ajoutée. La valorisation des données peut se voir limitée par leur format si celui-ci n'est pas conforme aux standards. La diversification des sources de données (capteurs, *crowdsourcing*) pose également des difficultés techniques. Enfin, pour les méga-données (*big data*) ou les données de temps réel, des questions de capacités techniques liées au réseau et au stockage peuvent se poser.

§ 3.3. L'exploitation des données est soumise à la résolution de *questions juridiques* souvent complexes. Les droits de propriété, droits d'utilisation et les « licences » qui les régissent ne sont pas toujours très clairs. Différents « secrets » peuvent également grever ces droits : secrets fiscal, médical, industriel et commercial, ou protection de la vie privée.

§ 3.4. Un dernier frein, souvent puissant, provient des incertitudes sur le partage de la valeur-ajoutée entre tous les acteurs qui contribuent à la production d'un nouveau service. Les *modèles économiques* des producteurs de données, comme l'ign, le mnhn ou les associations de surveillance de la qualité de l'air, s'opposent de fait à une circulation plus fluide des données qui conduirait à minorer leurs ressources propres. Les modèles de pseudo-gratuité de type Google ont par ailleurs répandu l'idée fautive que les données ne coûtaient rien à produire. De manière plus générale, la coopération que suppose la mise en partage des données doit s'articuler avec le contexte de concurrence économique des entreprises — par exemple dans le domaine du logement, du transport, de l'énergie, de l'eau ou des déchets.

### **§ 4. Il faut instituer une fonction, par la Ministre, pour accélérer la création de valeur promise par les données : le superviseur général des données du ministère.**

§ 4.1. Après un travail exploratoire avec les services du ministère et le réseau scientifique et technique, il s'avère que l'accélération de la valorisation de nos bases de données nécessite **l'institution d'une fonction de haut niveau, auprès de la Ministre**. Le superviseur des données lui rendra compte, par un *rapport annuel*, de l'avancée de la politique de la donnée du

ministère. L'attribution de cette fonction au Commissariat général au développement durable se justifie par l'expertise détenue dans ce domaine par son service de production des statistiques (soes) et sa mission d'information géographique (mig), ainsi que par sa mission d'animation du réseau scientifique et technique (rst). Cette fonction participe et bénéficie du réseau de l'Administrateur général des données, fonction rattachée au Premier Ministre. Un emploi chargé de la supervision générale des données du ministère serait identifié auprès de la Commissaire générale au développement durable. Quatre missions principales lui seraient confiées.

§ 4.2. Sa première mission consiste à **développer la culture de la donnée au sein du ministère**. En appui à un *cycle d'interventions de haut niveau à concevoir avec le Secrétariat Général*, il s'agirait de lancer plusieurs *expérimentations concrètes* avec les directions générales et les opérateurs concernés, dans l'objectif d'éprouver collectivement le bénéfice direct d'une meilleure exploitation des bases de données existantes. La base des diagnostics de performance énergétique des logements, la base des permis de construire et la base des immatriculations enrichie par des données d'émission des véhicules pourraient être l'objet de ces premières expérimentations d'application des data-sciences dans le ministère.

§ 4.3. Sa deuxième mission vise à **favoriser la connaissance et la circulation des données, et à mettre en relation les producteurs des données et les utilisateurs** qui pourraient s'en servir pour créer de la valeur-ajoutée. Dans le cadre de la *Green Tech*, cette intermédiation prendrait la forme notamment de « *hackathons* », et d'animation de l'incubateur. De nombreux thèmes sont envisageables : risques naturels, finance verte, gestion des alertes environnementales, économie circulaire... L'organisation d'un hackathon est l'occasion pour recenser et libérer toutes les données disponibles qui pourraient sous-tendre de nouveaux services dans le champ considéré.

§ 4.4. Sa troisième mission consiste à **identifier et promouvoir les données de référence dans le champ du ministère**. Les données de référence désignent les données utilisées par une multitude d'acteurs publics et privés, et dont la qualité et la disponibilité sont critiques pour ces utilisations. Elles constituent de véritables *infrastructures de données*.

§ 4.5. Sa quatrième mission porte sur la **production de doctrine commune**, au sein du ministère et du réseau scientifique et technique, en lien avec l'administrateur général des données, pour ce qui concerne les licences, la stratégie de plateforme, les modèles économiques et le partage de la création de valeur entre les partenaires, le *crowdsourcing*...

## **§ 5. Prochaines étapes**

§ 5.1. Création effective de la fonction : décret ministériel avant l'été 2016.

§ 5.2. Mise en place du réseau du superviseur de données avec l'ensemble des directions du ministère à l'automne 2016.

## Table des matières

<b>1. Beaucoup d'enjeux se nouent autour de la donnée pour l'action du MEEM.....</b>	<b>7</b>
1.1. Le MEEM joue un rôle unique dans la production de données en France.....	7
1.2. Des freins techniques limitent la circulation et l'usage des données .....	9
1.3. Des freins budgétaires et juridiques limitent la circulation et l'usage des données .....	10
1.4. Les secteurs économiques du MEEM subissent une profonde transformation par la donnée...11	
1.5. Le MEEM et la société civile tireraient quatre bénéfices à améliorer la circulation et l'usage des données .....	11
1.6. L'institution d'une fonction de superviseur général des données permettrait au MEEM d'accélérer collectivement.....	12
<b>2. Six voies sont ouvertes pour valoriser les données au MEEM .....</b>	<b>14</b>
2.1. Le superviseur général des données doit favoriser la montée en qualité des données .....	14
2.2. Le superviseur général des données doit poursuivre l'ouverture des données publiques du MEEM.....	14
2.3. Le superviseur général des données doit aider le MEEM à mieux utiliser ses propres données.....	16
2.4. Le superviseur général des données doit stimuler les écosystèmes qui utilisent les données du MEEM.....	17
2.5. Le superviseur général des données doit mettre les datasciences au service du MEEM .....	19
2.6. Le superviseur général des données est une fonction en réseau.....	19
<b>3. La fonction de superviseur général des données est une réponse concrète aux enjeux de la donnée.....</b>	<b>20</b>
3.1. Les objectifs .....	21
3.2. Les fonctions.....	21
3.3. Le programme de travail 2016/2017 .....	22
3.4. Les conditions de réussite et les indicateurs .....	23
3.5. Une proposition de rédaction réglementaire .....	24
<b>Annexe : lettre de mission .....</b>	<b>25</b>



## 1. Beaucoup d'enjeux se nouent autour de la donnée pour l'action du MEEM

La conduite de l'action publique engendre une *profusion de données*. Ces données sont de type et de qualités variés :

- des données de référence produites pour permettre à la collectivité nationale, européenne ou internationale de parler un langage commun,
- des données de gestion qui n'ont pas d'autre but initial que de rendre possible l'administration des politiques publiques,
- des données produites à des fins statistiques.

Pour autant, cet ensemble prodigieux de données est *sous-exploité*. La plupart des données peuvent s'avérer utiles bien au-delà du but pour lequel elles ont été élaborées en premier lieu. La démultiplication de l'utilité des données produites par les administrations publiques, en particulier au service de la transition, l'acculturation de l'administration et l'invention de nouveaux modes d'intervention de l'État sont les objectifs de la démarche de supervision des données que le ministère de l'environnement, de l'énergie et de la mer (MEEM) entreprend.

### 1.1. Le MEEM joue un rôle unique dans la production de données en France

Dans l'ensemble des ministères, le MEEM occupe une *place très particulière*. En effet, son champ de compétences, par essence territorial et environnemental, l'a conduit à mettre en place des opérateurs (par exemple Météo France et IGN) dont la mission est de produire de la donnée. Cette donnée de base est produite selon des exigences de qualité élevées qui font de ces produits une *véritable infrastructure collective*, c'est-à-dire un socle permettant le développement de l'activité de l'ensemble des acteurs de la société. Ce socle a pour double vertu de créer une norme permettant l'interopérabilité de productions particulières et de former un produit dans lequel tout le territoire national est traité de manière homogène. La cartographie de base ou la prévision nécessaire pour assurer la sécurité météorologique des personnes et des biens en sont des exemples.

Comme la plupart des autres ministères, le MEEM constitue également des bases de données à des fins *statistiques*. Pour cela il conduit des enquêtes ou traite des fichiers administratifs et met à disposition des chiffres de synthèse dont la qualité première est la cohérence statistique. Les bases de données sous-jacentes sont fiabilisées par des techniques classiques d'échantillonnage (sondage) qui permettent, sans nécessité d'atteindre l'exhaustivité, d'estimer des statistiques d'ensemble nécessaires à l'information, au débat et à l'évaluation de l'action publique. La mise à disposition

d'informations chiffrées et de leurs clés de lecture est partie intégrante de cette démarche qui répond désormais, le plus souvent, à des *obligations de rapportage fixées par la réglementation européenne*.

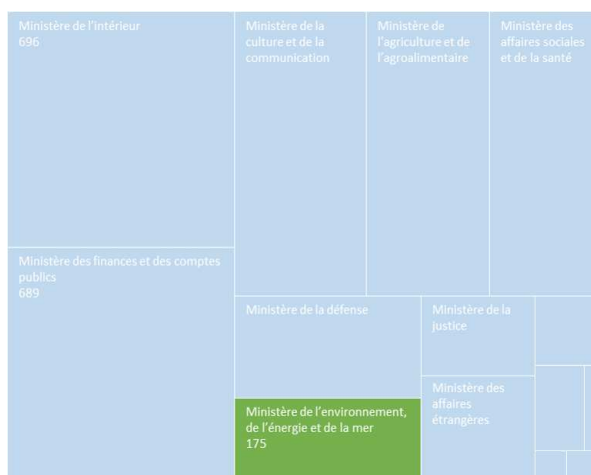
En parallèle, les services du MEEM et ses opérateurs ont développé, pour mettre en œuvre les politiques publiques dont ils ont la charge, des systèmes d'information qui organisent un très grand nombre de données. Certains établissements publics ont d'ailleurs reçu, parmi leurs missions, celle de produire et d'administrer des données — ainsi l'office national de l'eau et des milieux aquatiques (ONEMA), l'agence de l'environnement et de la maîtrise de l'énergie (ADEME), ou encore l'institut français de recherche pour l'exploitation de la mer (IFREMER)...

*La quantité de données produites par le MEEM est considérable* : pour la seule partie statistique, elle est estimée à 1 téraoctet, pour environ 100.000 sources de données. Les données accessibles par le Géoportail de l'IGN représentent un volume d'environ 250 téraoctets. Et, pour établir ses prévisions météorologiques, Météo France utilise un supercalculateur qui manipule plus de 1.500 téraoctets de données *par seconde*...

À titre de comparaison, les informations contenues dans la bibliothèque du Congrès américain, considérée comme la plus grande bibliothèque du monde avec ses 24 millions de livres, représentent un volume estimé à 18 téraoctets.

Une mission a été confiée, début 2016, au conseil général de l'environnement et du développement durable (CGEDD), pour établir une cartographie et une typologie des données dont dispose le MEEM et ses opérateurs ; son rapport est attendu pour septembre 2016.

Sur le portail interministériel d'accès aux données libres et ouvertes, [data.gouv.fr](http://data.gouv.fr), le MEEM est moins présent que d'autres ministères : ses données représentent 5 % des



ressources des ministères. Ce constat méritera d'être approfondi.

Au final, l'enjeu que représente la mise en valeur de ces données pour la collectivité est considérable. Leur ouverture et leur interopérabilité, insuffisamment abouties à ce jour, sont les conditions de cette mise en valeur.

## 1.2. Des freins techniques limitent la circulation et l'usage des données

Les données des bases de gestion ne sont pas produites pour des besoins publics de connaissance. Elles sont des sous-produits de l'activité d'administration des politiques publiques. Leur communication large est d'abord un gage de transparence. Elle est aussi un effort que consent par la collectivité : si la production de ces données est couverte par les financements publics, leur mise à disposition au-delà de leur usage initial constitue un service complémentaire. Il convient alors de documenter les bases de données diffusées et de les mettre sous un format intelligible par tous les acteurs susceptibles de les utiliser. Ces opérations ne suffisent pas : deux précautions doivent être prises en complément, concernant leur interopérabilité et leur qualité.

D'une part, ces données ont d'autant plus de valeur pour l'utilisateur qu'elles peuvent être aisément croisées avec d'autres sources et ceci suppose généralement qu'une clé permettant l'appariement ait été instaurée au sein des données. Par exemple, la base de gestion des permis de construire identifie les lieux à travers une adresse postale et un code cadastral déclarés. Pour faciliter l'*interopérabilité* de cette source, il convient de lui adjoindre une géolocalisation qui s'appuie sur les déclarations figurant dans le dossier, le cas échéant apurées par mise en cohérence avec des informations externes. Il arrive, en outre, que cette opération de référencement nécessite également la mise au point de la norme elle-même.

D'autre part, la *qualité des données* est évaluée relativement à leurs systèmes de gestion d'origine. Certains défauts d'exhaustivité, d'exactitude ou d'actualité, sans conséquence pour les buts assignés au système de gestion initial, peuvent s'avérer critiques pour l'usage de ces données à d'autres fins. La deuxième précaution à prendre, pour maximiser l'utilité collective des données, consiste donc à analyser les défauts des données par rapport à un schéma de qualité totale, à les porter à la connaissance des utilisateurs et le cas échéant, à faire en sorte que ces défauts soient atténués dans le cadre du processus de production. Cette analyse suppose aussi l'examen de la cohérence de la source concernée avec d'autres sources, notamment sur le plan statistique.

### **Remettre en accès des données déjà produites**

L'IFREMER a été chargé de développer un portail « système d'information scientifique pour la mer » (SISMER) qui compile, sauvegarde et diffuse des données collectées lors de programmes nationaux et internationaux.

L'IGN a collecté près de 4 millions de clichés aériens du territoire français dont les plus anciens remontent à 1920. Jusqu'en 2008, ces clichés étaient uniquement consultables à la photothèque nationale et reproduits sur demande dans le cadre d'un service payant. Depuis lors, l'IGN a engagé un vaste chantier de numérisation et de géolocalisation de ces clichés — désormais accessibles sur le géoportail et téléchargeables gratuitement.

### 1.3. Des freins budgétaires et juridiques limitent la circulation et l'usage des données

Rassembler des données pour assurer le suivi de mise en œuvre d'une politique publique, réaliser un référentiel normatif ou produire de l'information statistique est clairement une production de *bien public*. La collectivité supporte ainsi les coûts de cette production. En revanche, l'extraction, puis la transformation d'une donnée en donnée de diffusion est souvent un processus complexe mobilisant des compétences métiers et informatiques. La diffusion de ces données engendre alors un *coût de mise à disposition*. Or la loi pose le principe que la réutilisation de données publiques est *gratuite* — sauf exception, si l'équilibre budgétaire des opérateurs producteurs s'en trouvait affecté (cf. Loi du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public).

En matière d'information statistique, les règles européennes ont fixé le principe, depuis un peu plus d'une décennie, que l'information statistique, dans son ensemble, constitue un bien public. Pour la diffusion des données produites par les opérateurs dédiés, le modèle économique actuel faisant une part importante aux financements propres, le cadre reste à préciser au cas par cas. Cette réflexion doit naturellement s'inscrire dans une vision élargie à une politique ministérielle de la donnée.

D'autres freins peuvent également limiter la circulation et l'usage des données. Ces données peuvent être couvertes par des *secrets*. Ainsi, les bases de données contenant des informations personnelles n'ont pu être constituées qu'avec un engagement auprès de la commission nationale informatique et libertés (CNIL) détaillant les fins de leur utilisation. Changer l'usage de ses bases n'est donc pas possible sans préalable. Il faut souvent procéder à leur *anonymisation*, traitement difficile si l'on veut éviter tout risque de ré-identification. D'autres secrets peuvent être également opposés à la circulation des données : secret industriel et commercial, secret fiscal, secret médical...

D'une manière générale, les bases de données sont couvertes des droits de propriété intellectuelle et leur utilisation est encadrée par des licences.

#### Les données ouvertes, un bien public ?

Les « biens publics » sont non-rivaux et de non-exclusifs : non-rivalité signifie que la consommation du bien par un individu n'en réduit pas la quantité disponible pour les autres ; non-exclusivité signifie que le bien est disponible pour tous et que personne ne peut s'en arroger l'exclusivité.

La production de tels biens est rentable collectivement : leur valeur est supérieure à leur coût de production. Mais de tels biens ne sont pas spontanément produits par le marché. Ce type de bien et services doit en général être financé par l'impôt.

Pour les données publiques de référence, le bien produit est clairement un bien public : son caractère normatif et élémentaire rend possible d'autres productions qui sont nécessaires au développement d'une société moderne.

En revanche, sa diffusion peut-être exclusive, puisqu'il est possible de ne délivrer l'information qu'à certains acteurs. Avec l'avènement d'internet et pour les données numériques, le coût du service de mise à disposition devient négligeable. Il en découle que le processus d'ouverture des données s'impose naturellement pour des données produites en tant que bien public.

## 1.4. Les secteurs économiques du MEEM subissent une profonde transformation par la donnée

Le développement des flux d'information, avec une croissance forte pour ceux en temps réel, modifie considérablement l'administration des services (privés ou publics) rendus au public. L'offre privée est foisonnante et se construit sur les quantités extraordinaires de données individuelles accumulées par les propriétaires d'applications et de portails. Le rôle de régulateur que l'État exerce jusqu'à présent dans les services essentiels à la société bien que rendus par l'initiative privée s'en trouve potentiellement bouleversé et il convient d'examiner son positionnement stratégique à l'aune de cette révolution.

En particulier, s'il souhaite conserver un rôle de régulateur dans ce contexte, il doit s'adapter et agir avec les outils du numérique. La révolution numérique offre un avantage décisif à qui parvient, en tant que « premier entrant », à concentrer l'essentiel des échanges de données entre entreprises et/ou particuliers.

Le monopole de fait que détenait l'État sur certains échanges de données (géographie, transports, logements, énergie...) est aujourd'hui contesté par de grandes entreprises privées (Google, Uber, Airbnb, Tesla...). La mise en œuvre d'une politique de la donnée par l'État est donc porteuse de l'affirmation forte de son rôle de régulateur, sous le nom d'État-plateforme, stratégie à laquelle participe pleinement les services du ministère.

## 1.5. Le MEEM et la société civile tireraient plusieurs bénéfices à améliorer la circulation et l'usage des données

La richesse potentielle des bases de données se révèle sous plusieurs formes.

Une première valeur naît de l'ouverture au grand public. La transparence est nécessaire à la réussite de tout débat public, car elle contribue à la *confiance* et à l'*encapacitation* des participants, valeurs fondamentales. L'ouverture, prévue par la loi biodiversité, des données des études d'impact est ainsi très attendue. L'implication forte de la France dans le partenariat international pour un gouvernement ouvert (OGP) justifie un effort particulier dans ce domaine. La mise à disposition de ces données permettra également à l'ensemble de la société d'œuvrer à la transition écologique en inventant de nouveaux usages.

Une deuxième valeur naît lorsque les bases de données existantes sous-tendent la création de *nouveaux services* au public. Les données de gestion de la voirie d'une ville peuvent par exemple permettre de créer une application de choix d'itinéraires pour personnes à mobilité réduite. Les données d'inclinaison des toitures couplées aux données météo peuvent permettre d'anticiper le rendement de panneaux solaires pour les particuliers, comme l'expérimente la startup *In Sun We Trust*, incubée dans l'*IGNfab*. La valeur de ces services peut être marchande ou pas. Elles peuvent être créatrices d'emploi. Un des enjeux intimement lié au développement de cette valeur

est d'identifier les mécanismes économiques qui permettent de fonder le financement de l'ouverture des données sur les bénéfices collectifs avérés.

Que chacun puisse accéder à la donnée dont il a besoin pour exercer sa mission permet également des *économies* et des *gains de productivité*. Au-delà des bénéfices d'un meilleur accès des agents du ministère à la donnée utile, il est aussi possible de mieux allouer les moyens de l'État en s'appuyant sur de nouvelles analyses de données. Ainsi, des modèles prédictifs fondés sur les bases de données décrivant les installations classées et le résultat des contrôles passés peuvent permettre aux agents de l'État de cibler leurs contrôles sur les sites les plus susceptibles d'être en infraction. L'étude fine des consommations électriques des bâtiments administratifs peut permettre de renégocier les contrats d'énergie ou de prioriser les travaux de rénovation, comme l'a prouvé le hackathon *#CompteurConnect* (20-22 mai 2016).

Enfin, comme souligné au § 1.4., une valeur particulière pour l'État concerne son *positionnement stratégique* dans un environnement économique qui n'hésite plus à remettre en cause les prérogatives de puissance publique.

## **1.6. L'institution d'une fonction de superviseur général des données permettrait au MEEM d'accélérer collectivement**

Améliorer la circulation et l'usage des données du MEEM pour accompagner la transition énergétique et écologique passe par une *politique de la donnée*, véritable mouvement d'ensemble du ministère.

Beaucoup d'initiatives sont déjà en cours dans le domaine du numérique, notamment au secrétariat général (SG), au commissariat général au développement durable (CGDD), dans les directions générales (DG), ainsi que dans les opérateurs du réseau scientifique et technique (RST). La Green Tech – technologies vertes a déjà à son actif plusieurs événements autour des données du ministère.

Toutes les problématiques numériques ne se résument pas à une question de circulation des données, mais une coordination étroite est nécessaire entre la transformation numérique du ministère, promue par le SG, et la politique de la donnée.

Une politique de la donnée consiste à organiser les efforts du ministère pour améliorer la circulation et l'usage des données. Elle prend la forme d'un programme transverse au ministère. Elle stimule, coordonne et priorise les différents chantiers dans le domaine de la donnée.

La création d'une fonction de superviseur général des données doit précéder l'élaboration de la politique de la donnée du ministère.

Son périmètre d'action doit être large. La circulation et l'usage des données concernent tout autant les échanges à l'intérieur des administrations centrales et déconcentrées du MEEM que l'ouverture externe, y compris au grand public. Les données concernées ne sont pas seulement celles qui sont produites par le MEEM et

ses opérateurs, mais aussi celles qui sont produites par d'autres (collectivités, entreprises délégataires du service public...) et qui sont utiles à la régulation.

La nature très diverse de ces données — données de référence comme la Base Adresse Nationale, données de gestion comme les consommations énergétiques des ménages, données statistiques comme les diagnostics de performance énergétique — se reflétera dans le portefeuille des projets suivis.

La fonction de superviseur général des données, au-delà de son rôle stratégique et d'entraînement collectif, comprend également la mise en place d'une force d'appui destinée à enrichir les sources, les qualifier et à développer des innovations autour des données.

## 2. Six voies sont ouvertes pour valoriser les données au MEEM

Le superviseur général des données du MEEM, en charge de la politique de la donnée, a pour objectif principal d'augmenter la valeur des données, selon les différentes acceptions de ce terme précisées au 1.5. Il dispose de six voies principales pour cela.

### 2.1. Le superviseur général des données doit favoriser la montée en qualité des données

Une qualité élevée des données rend possible des traitements élaborés de modélisation ou de prédiction. Par nature, cette exigence de mise en qualité des données est une des activités premières des métiers de la statistique publique. Elle consiste, en amont des analyses statistiques, en une expertise des données. Celle-ci vise à évaluer la qualité des données, leur cohérence et, si besoin identifier les actions possibles pour les améliorer. Cette bonne pratique d'évaluation de la qualité est un gage de maîtrise et de maturité des données produites et détenues.

Les services du ministère contribuent à l'évaluation des données collectées ou détenues, par exemple au travers des données statistiques produites lors des enquêtes, ou encore à la demande des partenaires, comme pour les diagnostics de performance énergétique.

Par ailleurs, la mise en qualité au travers de procédures statistiques n'est pas la seule voie possible. Pour des données d'observations par exemple, il existe des solutions alternatives grâce à la mutualisation et à la mobilisation citoyenne. Par exemple, le projet Base Adresse Nationale (BAN) a mis en commun les données de l'IGN, de la Poste, du cadastre et des données produites par les contributeurs d'OpenStreetMap. Les données d'adresses sont ainsi améliorées en continue grâce à une contribution de partenaires institutionnels et de contributions citoyennes.

#### La base des DPE

Le diagnostic de performance énergétique (DPE) renseigne la performance énergétique d'un logement ou d'un bâtiment, en évaluant sa consommation d'énergie et son impact en termes d'émission de gaz à effet de serre.

Par la loi de 2010 portant engagement national pour l'environnement, les DPE doivent être transmis à l'ADEME.

Le service statistique du ministère a réalisé une analyse de qualité des données des DPE. Sur la base de règles de contrôle simples, il apparaît que près de 20 % des données disponibles dans la base des DPE sont des données redondantes. En associant les entreprises de diagnostics au processus de remontée des informations, une meilleure gestion des données permettrait d'augmenter la qualité des données diffusées. À défaut, une procédure de mise en qualité est relativement simple à industrialiser dans le processus de diffusion de l'information sur le site de l'ADEME.

### 2.2. Le superviseur général des données doit poursuivre



## L'ouverture des données publiques du MEEM

L'ouverture des données publiques est une exigence démocratique. Son principe est affirmé par les directives européennes et par la loi. En facilitant l'accès et la réutilisation de données produites dans le cadre des missions du service public, elle permet d'améliorer la transparence sur l'action de l'État. Elle offre également des ressources au développement d'activités, aussi bien publiques que privées. Le ministère de l'environnement, de l'énergie et de la mer (MEEM) s'est engagé résolument dans cette voie depuis plusieurs années. Plusieurs initiatives nouvelles sont en cours.

### Les permis de construire

Les informations contenues dans les formulaires de permis de construire de logements et de locaux non résidentiels constituent la source des statistiques de la construction. Ces informations rassemblées dans la base de données « Sit@del2 » permettent le suivi conjoncturel et une analyse structurelle de la construction.

#### **Plateforme Géorisques**

Depuis la mise en place de la directive Inspire, la direction générale de la prévention des risques (DGPR) met à disposition des jeux de données « risques ». Ces données sont produites en partenariat avec les services déconcentrés et les établissements publics. La plateforme Georisques diffuse 170 couches de données vers le grand public et gère la diffusion de 220 couches vers l'administration. Ce sont près de 800 000 connexions qui sont enregistrées tous les ans et 4 documents édités par minute sur les risques pour un point précis du territoire.

Une offre de service en *opendata* est en cours de développement. Elle inclurait, en standard, la mise à disposition de fichiers déclinés selon plusieurs critères d'intérêt économique : type de construction, statut du permis de construire (déposé, autorisé, commencé) et champ géographique (département, région ou France entière).

L'ouverture de ces données impose au préalable une réflexion juridique aboutie. Même si elles sont anonymes, il faut s'assurer que ces données ne puissent, par recoupement avec d'autres, enfreindre la protection de la vie privée. Cet exemple illustre la nécessité de fonder l'ouverture des données sur une doctrine éprouvée des enjeux d'identification indirecte des personnes.

### Les données routières

Le MEEM a ouvert, début 2016, ses données sur le domaine du réseau routier national. Ce nouvel accès aux données de réseau participe à l'amélioration de la sécurité routière des usagers. Les décideurs et les citoyens sont informés en toute transparence sur l'état du patrimoine, l'activité du réseau et l'utilisation de l'argent public. La mise en ligne de ces données illustre la volonté d'un partage le plus large possible de la connaissance des actions menées sur les infrastructures routières par les services du ministère. Au-delà de la transparence, l'impact recherché est double : diminuer le nombre d'accidents et réduire leur gravité.

En septembre 2016, les premières données statistiques seront publiées : consistance du

réseau, état du patrimoine, activités des gestionnaires et dépenses d'entretien. Dans un an, de nouvelles données seront ajoutées *en temps réel* : accidents, travaux, fermeture des voies, volume du trafic, vitesse, longueur des bouchons. Toutes les données seront diffusées gratuitement et réutilisables par la communauté numérique.

### **2.3. Le superviseur général des données doit aider le MEEM à mieux utiliser ses propres données**

De nombreuses missions assumées par les administrations du MEEM peuvent gagner en efficacité grâce à un usage pertinent des données déjà disponibles. En particulier, le potentiel d'aide à la décision de l'information géographique est encore souvent sous-estimé : les données sont plus souvent utilisées pour des représentations cartographiques que pour des analyses proprement dites.

Définir les zones favorables pour l'implantation d'un projet, repérer les signaux faibles territoriaux annonciateurs de difficultés socio-économiques, mais aussi choisir les zones à contrôler en priorité : ces activités habituelles dans les services du MEEM peuvent bénéficier d'une meilleure utilisation des données.

#### **L'importance des données de gestion**

À la différence des données qui ont été pensées pour décrire un objet — par exemple la cartographie d'une route — les données de gestion désignent les informations produites par le gestionnaire de l'objet pour ses propres missions — par exemple la surveillance du trafic routier pour éviter les accidents. Les systèmes de gestion recèlent souvent des informations très précieuses pour décrire les systèmes de manière dynamique, et même prédictive.

Or, souvent, les données de gestion ne sont pas considérées comme des données à proprement parler. Il n'est pas toujours facile de les extraire de leurs systèmes d'information ou de les homogénéiser. La première étape est de mieux les intégrer dans les travaux de l'administration et d'en assurer une pleine appropriation. Elle s'appuie sur la capacité des services — capacité technique voire financière lorsqu'ils sont sous-traités à ces sociétés sous-traitantes — à faire évoluer ces systèmes de gestion, souvent jugés complexes.

La direction générale de la prévention des risques (DGPR) assure un suivi des installations classées. Un système de gestion permet à chaque inspecteur d'enregistrer les contrôles qu'il a réalisés. Après une phase d'amélioration de l'usage du système de gestion et la fiabilisation des données enregistrées, la DGPR a constaté que le système de gestion recelait un « trésor » : la liste des contrôles réalisés permet d'élaborer une cartographie des contrôles les plus efficaces, en fonction du type d'installations classées. Dans un contexte où la meilleure allocation des agents sur le terrain est cruciale, une stratégie a été développée pour faire porter l'effort de contrôle en priorité les installations les plus sensibles.

Ainsi, la direction générale de la prévention des risques (DGPR) utilise ses bases de données de gestions des contrôles d'installations classées pour mieux orienter son effort en la matière.

Autre exemple : dans le cadre de la politique « État exemplaire », pour une maîtrise des émissions de gaz à effet de serre, le secrétariat général du ministère établit le bilan carbone du ministère.

C'est au travers des marchés publics de voyagistes passés par le ministère qu'il a pu accéder aux données de déplacements professionnels des agents. Une fois consolidées avec l'ensemble des services, ces données de gestion, normalement destinées à payer les transporteurs, permettent de dresser un panorama complet des kilomètres parcourus et des émissions de gaz à effet de serre liés aux déplacements professionnels.

**Émissions de CO<sub>2</sub>** dues aux déplacements professionnels des agents du CGDD (en teq CO<sub>2</sub>) :

2013	2014	2015
87,1	90,1	90,3

Malgré les évolutions techniques qui ont permis la réduction du CO<sub>2</sub> par km, le bilan carbone des déplacements professionnels s'est alourdi de 3,6 % entre 2013 et 2015. En particulier, le recours à des déplacements par avion est plus fréquent.

Une réflexion approfondie sur les déplacements professionnels et la compensation des gaz à effet de serre est donc nécessaire pour endiguer cette évolution.

## 2.4. Le superviseur général des données doit stimuler les écosystèmes qui utilisent les données du MEEM

Selon la formule consacrée, « il y a toujours plus d'intelligence à l'extérieur d'une organisation qu'à l'intérieur ». C'est l'un des enjeux actuels, pour l'État comme pour tous les acteurs économiques : comment attirer, comment mobiliser, comment fidéliser toute cette « intelligence extérieure » pour prolonger son action ?

Pour la constitution de tels écosystèmes, le MEEM dispose d'avantages décisifs : la masse et la qualité de ses données, l'importance des secteurs économiques qu'il couvre (logement, énergie, transport, risques, aménagement, déchets, eau, pêche...), ainsi que le militantisme d'associations environnementales (qualité de l'air, biodiversité...).

Une méthode standard pour aller à la rencontre de ses écosystèmes passe par l'organisation de hackathons. Ces événements consistent à mettre en partage des données auprès d'équipes bénévoles, qui vont en 24 ou 48h élaborer la preuve de concept d'un nouveau service numérique utilisant les données.

Pour que la rencontre soit fructueuse, elle doit être préparée minutieusement, en amont, par un travail sur les données (nouveau, montée en qualité, documentation, anonymisation...) ainsi que par l'élaboration d'une liste de « problèmes à résoudre » pour améliorer la vie des citoyens, pour faciliter le travail des entreprises, pour mettre en valeur le cadre de vie...

Deux mois avant le hackathon, une journée d'intelligence collective ouverte à tous est organisée. Des ateliers auto-administrés listent les enjeux et les pistes de défi à relever par les développeurs. Il est important de rassembler des personnes de tous horizons : étudiants, entrepreneurs, citoyens engagés, associations, grandes entreprises...

En parallèle, un travail renforcé avec les producteurs de données — les DG, les opérateurs du ministère, les entreprises publiques, voire des entreprises privées — permet d'obtenir les agréments nécessaires pour mettre en partage des données attractives.

Il est à noter que les données partagées lors du hackathon ne relèvent pas forcément de l'*opendata*. Par exemple, lors du hackathon *#CompteurConnect* du 20 au 22 mai 2016, les données mises à disposition par ErDF et GrDF n'étaient accessibles qu'aux participants et que pendant la durée de l'événement.

De même, une partie des 35 millions de données rendues accessibles par le muséum national d'histoire naturelle (MNHN) lors du hackathon *#HackBiodiv* du 3 au 5 juin 2016 ne le seront que pour l'événement.

Les résultats visibles et directs des hackathons sont la démonstration des possibilités de création de nouveaux services sur la base des données mises en partage. Ces ébauches de services peuvent ensuite donner lieu à des développements plus aboutis, voire à la création d'entreprises, qu'il sera possible d'incuber dans le cadre de la Green Tech – technologies vertes.

L'intérêt principal pour le MEEM est de se constituer des communautés d'utilisateurs de ses données. Le superviseur général des données veillera à ce qu'elles soient animées dans la durée.

Un aspect particulier de ces écosystèmes porte sur la construction collaborative de bases de données : le *crowdsourcing*. Sur le modèle de Wikipedia ou d'OpenStreetMap, il s'agit d'organiser la contribution de producteurs de données bénévoles. Le MNHN a constitué une grande partie de ses bases de données ainsi. Un autre exemple porte sur la Base Adresse Nationale, déjà mentionnée.

Avec la mise en place de la GreenTech verte, le ministère lance un appel à projets dans huit secteurs clés pour placer le numérique au service des citoyens : les économies d'énergie, les énergies renouvelables, le bâtiment, le transport, les risques, la santé et l'environnement, la biodiversité.

Au sein du ministère, un incubateur GreenTech verte sera ouvert aux jeunes pousses dont les projets innovants participent à la transition écologique. 50 projets seront sélectionnés.

Des hackathons seront organisés sur ces thèmes.

Enfin, un concours est organisé dans les écoles du ministère pour sélectionner 2 projets portés par des élèves.

## **2.5. Le superviseur général des données doit mettre les datasciences au service du MEEM**

Grâce au marché d'appui proposé par l'Administrateur général des données, le superviseur contribuera à soutenir les initiatives des services en organisant le partage des expériences et la mise en réseau des connaissances. L'objectif est triple : créer une communauté d'experts sur les outils de datascience, stimuler la montée en compétence et impulser le partage des bonnes pratiques.

Le superviseur a pour ambition de stimuler le meilleur usage des données par l'administration grâce à des méthodes de « datascience ». Trois axes peuvent être encouragés dans ce domaine : la recherche et l'exploration des données ; l'analyse des données, les modèles prédictifs, la détection de signaux faibles, la classification, le « machine learning » ; la restitution de données.

Deux actions ont déjà été engagées début 2016. La première concerne l'expérimentation d'une plate-forme de traitements en masse des données. Cette plate-forme permettra un accès complet aux données pour les statisticiens du ministère, soit près de 1 teraoctet d'informations structurées, classées et documentées. La plate-forme permettra de croiser des données riches et volumineuses, mais également de réaliser des analyses complexes nécessitant des capacités de calcul élevées et d'élaborer des restitutions dynamiques.

La seconde action concerne l'accompagnement des agents du ministère aux outils et solutions de traitement des données, en particulier sur les outils open source. Grâce à l'association des connaissances en DREAL et du pôle national d'expertise « logiciels statistiques » hébergé par le service statistique du ministère (SOeS), un groupe de référents du logiciel statistique gratuit R élabore un cursus de formation pédagogique et progressif à l'outil. Les premières séances auront lieu début juin 2016. La formation pourrait être proposée début 2017 sous un format « formation à distance ».

## **2.6. Le superviseur général des données est une fonction en réseau**

On l'a vu, de nombreuses initiatives autour de la donnée sont déjà lancées au MEEM. Il est utile de les stimuler et de les coordonner pour en tirer le meilleur impact. À cet effet, il est proposé que le superviseur général des données puisse animer un réseau de correspondants dans les directions générales et dans les opérateurs du réseau scientifique et technique du ministère.

De même, le superviseur général des données s'inscrit dans le réseau interministériel animé par l'Administrateur général des données rattaché au Premier ministre.

### 3. La fonction de superviseur général des données est une réponse concrète aux enjeux de la donnée

Les éléments qui précèdent justifient l'intérêt d'instaurer une mission de supervision des données du ministère chargée de définir et de conduire une politique ministérielle de la donnée. Elle serait indubitablement un apport conséquent au service de la transition, de la modernisation de l'administration du ministère et de l'encapacitation des acteurs.

Une telle mission devra nécessairement être fortement articulée avec l'administrateur général des données et devra s'inscrire comme une sorte de relais ministériel de l'ambition portée par cette institution. En marge des attributions du superviseur des données, l'utilité d'instaurer une gouvernance des données ministérielles s'est faite sentir, pour les mêmes raisons que celles avancées dans ce rapport, au ministère de l'emploi et de la solidarité. Ainsi, la direction de la recherche, des études, de l'évaluation et des statistiques en matière de santé et de solidarité du ministère des affaires sociales et de la santé (DREES) a été chargée, en 2015, d'exercer, pour le compte de son ministère la gouvernance des données de santé. Le décret du 21 juillet 2000 relatif à l'organisation de l'administration centrale du ministère de l'emploi et de la solidarité a été complété d'un nouvel alinéa, concernant les attributions de la DREES, ainsi rédigé : « *En liaison avec l'administrateur général des données, la direction est chargée de la gouvernance des données de santé et de la définition des règles de leur mise à disposition à des fins de connaissance, d'étude, de recherche et d'évaluation.* »

Il est donc proposé, sur l'exemple du ministère de l'emploi et de la solidarité, d'instituer par décret, auprès de la ministre de l'environnement, de l'énergie et de la mer, un superviseur général des données (SGD). Le superviseur général des données du ministère anime et coordonne l'action des services, tels que définis dans le décret 9 juillet 2008 portant organisation de l'administration centrale du ministère, et conformément au décret d'attribution de la ministre, en matière d'inventaire, de gouvernance, de production, de circulation et d'exploitation des données.

De manière collective et collaborative avec les directions générales, et des opérateurs sous tutelle, il organise, dans le respect de la protection des données personnelles et des secrets protégés par la loi, la meilleure exploitation de ces données et leur plus large circulation, notamment aux fins d'évaluation des politiques publiques, d'amélioration et de transparence de l'action publique et de stimulation de la recherche et de l'innovation.

L'attribution de cette fonction au Commissariat général au développement durable se justifie par la capacité de son service d'observation et des statistiques (SOES) et de sa mission d'information géographique (MIG) ainsi que par sa mission d'animation du réseau scientifique et technique (RST).

La fonction de superviseur général des données a donc vocation à être créée auprès de la Commissaire générale au développement durable. Le titulaire de la fonction devra témoigner d'une expertise avérée du sujet et d'une grande expérience du ministère qui lui permette de se placer comme interlocuteur légitime du secrétaire général et des directeurs généraux du ministère.

### 3.1. Les objectifs

La mission de supervision recouvre les objectifs fixés dans la lettre de mission de la ministre de l'environnement de l'énergie et de la mer à la commissaire générale au développement durable (annexe). Ces objectifs sont :

- d'organiser, la meilleure exploitation de données et leur plus large circulation, notamment aux fins d'évaluation des politiques publiques, d'amélioration et de transparence de l'action publique et de stimulation de la recherche et de l'innovation;
- de coordonner l'action des directions et des opérateurs en termes d'inventaire et de gouvernance de production des données ; cette coordination comprend la participation aux relations qu'entretient le ministère avec les autorités administratives indépendantes s'agissant des données du secteur concerné.
- de proposer des stratégies d'exploitation des données produites, reçues ou collectées par l'ensemble des deux ministères et des opérateurs dans le cadre de leurs missions de service public, y compris en s'appuyant sur des entreprises innovantes ;
- d'élaborer des outils, des référentiels et des méthodologies permettant une meilleure exploitation des données et un plus grand usage des sciences des données;
- de conduire des expérimentations sur l'utilisation des données pour renforcer l'efficacité des politiques publiques, contribuer à la bonne gestion des deniers publics et améliorer la qualité des services rendus aux usagers.

### 3.2. Les fonctions

Dans le détail, le superviseur des données assure les fonctions suivantes :

- Gouvernance des données
  - Définir la politique de la donnée pour le ministère ;
  - Accompagner les services du ministère dans leur réflexion en matière de régulation des secteurs économiques de leur compétence, dans le contexte du développement exponentiel des services de la donnée ;
  - Coordonner l'action des directions et des opérateurs en termes

d'inventaire, de production et de diffusion des données et des données de référence ;

- Assurer une veille sur les textes juridiques, en particulier la protection de la vie privée, impactant les données et leur ouverture ;
- Exploitation et circulation des données
  - Déterminer des stratégies d'exploitation des données produites, reçues ou collectées en mobilisant les services du ministère (DG et opérateurs) dans le cadre de leurs missions de service public ;
  - Expertiser, avec les directions et les opérateurs, les impacts sur les modèles économiques et le partage de la création de valeur sur la donnée ;
  - Organiser la meilleure exploitation de données et leur plus large circulation, notamment aux fins d'évaluation des politiques publiques, d'amélioration et de transparence de l'action publique et de stimulation de la recherche et de l'innovation ;
    - Assister les services sur les clauses des marchés publics en cas de recours à la sous-traitance ;
    - Renforcer le lien entre les maîtrises d'ouvrage et les maîtrises d'œuvre des systèmes d'information (en particulier avec CPII)
- Outils et expérimentations
  - Élaborer des outils (de type guichet unique des données), des référentiels et des méthodologies permettant une meilleure exploitation des données et un plus grand usage des sciences des données ;
  - Conduire des expérimentations sur l'utilisation des données pour renforcer l'efficacité des politiques publiques, contribuer à la bonne gestion des deniers publics et améliorer la qualité des services rendus aux usagers.

### 3.3. Le programme de travail 2016/2017

- Gouvernance des données
  - Définir, mettre en place et animer un comité stratégique inter-services sur la gouvernance des données
  - Stimuler et coordonner les initiatives prises par les directions générales et les opérateurs sur la gouvernance des données
  - Apporter un appui technique ou méthodologique aux travaux de gouvernance des données
  - Assurer la coordination des travaux du ministère avec le programme de travail de l'administrateur général des données
  - Réaliser un inventaire des données, définir un processus de mise à jour annuelle



- Définir les données de référence sous la responsabilité du ministère, et animer leur mise à disposition
- Rédiger une synthèse sur les textes juridiques, dont la protection de la vie privée, en vigueur et à venir, impactant les données et leur ouverture
- Rédiger un premier rapport annuel 2017
- Participer aux actions du ministère retenues dans le cadre de l'OGP (décembre)
- Exploitation et circulation des données
  - Organiser des hackathons sur les données du ministère
  - Mettre en place un processus de vérification des appels d'offre ou commandes pour des travaux de construction « SI » supérieur à un certain montant
- Outils et expérimentations
  - Construire un silo de stockage des données énergie (150 opérateurs d'ici la fin 2016)
  - Créer une offre de formation des agents au logiciel libre de traitements statistiques
  - Expertises juridique et technique pour une ouverture des données des permis de construire (Sitadel) et des immatriculations de véhicules (en lien avec le ministère de l'intérieur)
  - Séminaire DG sur la culture de transition ; favoriser la diffusion de la culture de la donnée au sein du RST

### 3.4. Les conditions de réussite et les indicateurs

La réussite de la mission des données repose sur plusieurs facteurs :

- une participation active du secrétariat général, des directions générales et des opérateurs
- des objectifs ambitieux et réalistes
- une équipe resserrée et pluridisciplinaire (statisticiens, informaticiens, experts métiers, infographistes)
- des moyens d'investissement sur des outils ou nouvelles technologies

Pour évaluer les gains obtenus par les services du ministère et des opérateurs grâce à la supervision des données, une série d'indicateurs sera mise en place au cours du 4<sup>e</sup> trimestre 2016. Ces indicateurs, simples à mesurer, permettront d'illustrer le chemin accompli par les services du ministère dans les différentes actions de la mission de

supervision : connaissance, partage et réutilisation des données, mise en œuvre de bonnes pratiques, expertises juridiques, pistes sur le partage de la création de valeur, réalisation de projets innovants, développement des datasciences, etc.

### 3.5. Une proposition de rédaction réglementaire

L'inscription de la fonction de supervision générale des données du ministère se traduit par le projet de décret suivant (modification du décret du 9 juillet 2008) :

*« En cohérence avec la politique du gouvernement mise en œuvre par le secrétariat général pour la modernisation de l'action publique et en appui aux directions générales, le commissariat général au développement durable est chargé de la supervision générale des données sur le logement, l'énergie, les transports et l'environnement, de l'évaluation et de la mobilisation des moyens nécessaires pour le développement de leur mise à disposition à des fins de connaissance, d'étude, de recherche et d'évaluation. »*

Les missions de la DRI et du SOeS sont par ailleurs complétées ainsi :

*« Au titre de la supervision générale des données, la direction de la recherche et de l'innovation est chargée de l'animation de l'activité économique, de la coordination inter-services relevant de son périmètre et de la définition d'une politique de la donnée pour les opérateurs dont elle assure la tutelle. »*

*« A ce titre et dans le cadre de la supervision générale des données, le service de l'observation et des statistiques est en outre chargé de la qualification, du traitement et de la diffusion de la donnée ainsi que de l'appui aux services du ministère dans ce domaine ; il développe le savoir-faire du ministère en matière de traitement et de sciences des données. »*

L'arrêté (9 juillet 2008) portant organisation du ministère devra être modifié pour inscrire le fonctionnement de cette nouvelle fonction dans l'organisation du CGDD. Celle-ci s'appuie sur les compétences en matière de données, d'analyse statistique et de systèmes d'informations ainsi que sur les capacités d'innovation et d'animation de la recherche actuellement détenues par le Service de l'Observation et des Statistiques (SOeS) et la Direction de la Recherche et de l'Innovation (DRI). Il convient de faire en sorte que ces compétences puissent être efficacement mobilisées pour les besoins de la supervision générale des données et qu'elles se renforcent. C'est un des buts poursuivis par le projet de service du CGDD actuellement en cours d'élaboration.

## Annexe : lettre de mission



*La ministre de l'Environnement,  
de l'Énergie et de la Mer,  
chargée des relations internationales sur le Climat*

*Ségolène Royal*

Paris, le 18 FEV. 2016

Objet : Superviseur des données

Madame la Commissaire générale,

Consommations d'énergie, performance énergétique des bâtiments, état du parc automobile, circulation, risques naturels et industriels, biodiversité, permis de construire, météorologie... mais aussi référentiels pour l'information géographique ou modèles de prix de l'énergie... Le ministère et son réseau scientifique et technique disposent d'une immense richesse de données, tant par sa diversité que par sa profondeur.

Cette richesse est actuellement sous-utilisée. D'abord, parce qu'il faut compléter la cartographie de ces bases de données, qui ont été développées sans cadre commun préalable, pour des objectifs de politiques publiques précis et non dans une perspective ouverte à d'autres usages. Ensuite, pour des raisons techniques ou juridiques posées par l'interopérabilité et la circulation des données. Enfin, parce que les usages nouveaux de ces données ne peuvent être inventés qu'en rassemblant ceux qui les produisent, ceux qui les agrègent, ceux qui s'en servent, et d'autres, qui ne s'en servent pas encore.

En lien avec la fonction d'administrateur général des données qui a été créée en septembre 2014 au sein des services du premier ministre, et sur son modèle, la fonction de superviseur des données (« *data officer* ») du ministère aura pour rôle :

**Madame Laurence MONNOYER-SMITH**  
Commissaire générale au Développement durable  
Tour Séquoia  
92055 La Défense

*Hôtel de Roquelaure - 246, Boulevard Saint-Germain - 75007 Paris*

- D'organiser la meilleure exploitation de données et leur plus large circulation, notamment aux fins d'évaluation des politiques publiques, d'amélioration et de transparence de l'action publique et de stimulation de la recherche et de l'innovation ;
- De coordonner l'action des directions et des opérateurs en termes d'inventaire et de gouvernance de production des données ;
- De proposer des stratégies d'exploitation des données produites, reçues ou collectées par l'ensemble des directions et des opérateurs dans le cadre de leurs missions de service public, y compris en s'appuyant sur des entreprises innovantes ;
- D'élaborer des outils, des référentiels et des méthodologies permettant une meilleure exploitation des données et un plus grand usage des sciences des données ;
- De conduire des expérimentations sur l'utilisation des données pour renforcer l'efficacité des politiques publiques, contribuer à la bonne gestion des deniers publics et améliorer la qualité des services rendus aux usagers.

Vous me proposerez, au plus tard le 31 mai prochain, un rapport de préfiguration du rôle de superviseur. Vous veillerez à préciser les scénarios en termes d'organisation de façon à mettre en œuvre une fonction de superviseur des données au sein du ministère et opérateurs sous tutelle à l'horizon de septembre de cette même année.

Je vous prie de croire, Madame la Commissaire générale, en l'assurance de mes sentiments les meilleurs.

  
Ségolène ROYAL





*Liberté • Égalité • Fraternité*  
RÉPUBLIQUE FRANÇAISE

MINISTÈRE  
DE L'ENVIRONNEMENT,  
DE L'ÉNERGIE  
ET DE LA MER