



**MINISTÈRE  
DE L'ÉCONOMIE,  
DES FINANCES  
ET DE LA RELANCE**

*Liberté  
Égalité  
Fraternité*

**Secrétariat  
général**

# Faire de la donnée un actif stratégique du MEFR

Feuille de route 2020-21

Août 2020



## Préface

Par son caractère imprévisible et incontournable, la crise sanitaire a chamboulé en quelques jours nos missions, nos habitudes de travail, nos priorités. Il est encore trop tôt pour tirer un bilan objectif, mais ce bouleversement inédit est révélateur de nos actions, comme accompagner les entreprises dans leur développement ou lors de leurs difficultés, fluidifier les échanges de marchandises avec les pays étrangers, détecter les fraudes ou encore produire les indicateurs statistiques et économiques. Or la donnée est toujours plus au cœur de toutes ces actions pour piloter nos actions et aider à prendre les bonnes décisions. Une plus forte utilisation des données devient alors aujourd'hui centrale pour gagner en efficacité et développer un service public performant.

La fonction d'administrateur ministériel des données (AMD) du MEFR a été créée en avril 2019 pour **accompagner résolument nos directions à la « révolution de la donnée »**, en permettant une meilleure mobilisation des données dans le pilotage des politiques publiques ou dans la création de nouveaux services numériques. Cette ambition exige la capacité à trouver et à exploiter les données, la bonne maîtrise des outils de la science de la donnée (analyse, modélisation, prédiction, intelligence artificielle), mais surtout la culture suffisante pour mobiliser et comprendre des méthodes parfois très complexes qui sont utilisées à des fins d'amélioration du service public ou du débat public.

Construit initialement dans un esprit de « startup d'État » avec une équipe réduite et avec une approche « produit » centrée vers les directions, la fonction AMD devait relever un défi majeur : créer les conditions favorables pour que le MEFR tende vers une meilleure maîtrise de ses données, les utilisent, les partagent – y compris dans le respect des secrets légaux et de la protection des données personnelles – pour concevoir et piloter les politiques publiques.

**Les premières actions lancées en 2019, dont l'installation du réseau des administrateurs directionnels de données, la journée de la donnée et les premières conférences-débats, ont permis de relever ce défi et ont démontré l'intérêt de cette fonction AMD**, et par conséquent l'intérêt des données dans toutes nos missions. La grande maturité des directions aux données à Bercy est également à souligner.

**Cette feuille de route 2020/2021 crée un fil d'Ariane, qui permet à toutes les directions de s'engager sur un chemin commun dans le grand labyrinthe numérique de la donnée.** Elle place les agents du MEFR au cœur de la transformation en misant sur le développement des connaissances et des enjeux. Elle appelle également à plusieurs investissements, comme le développement actif du service public de la donnée et les usages innovants de la donnée fortement soutenus par le fonds de transformation ministérielle (FTM).



Nous baignons d'ores et déjà dans le monde des algorithmes et de l'intelligence artificielle. Éloigné de la bataille industrielle engagée par plusieurs pays pour prendre la place de leader en IA, le MEFR doit se placer au niveau de la bataille culturelle et des usages de l'IA. Or sans donnée, l'intérêt de ces outils est moindre. Cette feuille de route pose donc 25 premières actions tactiques à l'échelle ministérielle pour y parvenir. La participation de toutes les directions, des agents et des managers, est cruciale pour positionner le MEFR dans le siècle des données.

**La donnée a vocation à nourrir la transformation et de la modernisation de l'État**, dans les territoires comme en centrale. Elle porte l'ambition d'une puissance publique opérationnelle et efficace au service de la République : « la démocratie, gouvernement du peuple, par le peuple, pour le peuple ».

La secrétaire générale du ministère de l'économie, des finances et de la relance  
Marie-Anne BARBAT-LAYANI



## Synthèse

Balance comptable des collectivités, cessions immobilières, nomenclature combinée des marchandises, projets de loi de finance, comptabilité générale de l'État, impôts locaux et octroi de mer, statistiques douanières, fiscales ou économiques... mais aussi référentiel des entreprises, code officiel géographique, nomenclature française d'activités... Le ministère dispose d'une immense richesse de données tant par sa diversité que par sa profondeur et son expérience.

**La donnée est au cœur de la transformation publique.** Des économies sur le fonctionnement à l'innovation des services numériques, son impact est réel dans toutes les activités pour qui sait l'exploiter. Elle aide les politiques ou les décideurs à prendre des décisions, les managers à piloter les missions, les agents à faciliter leurs activités, le grand public à exercer leurs droits de contrôle des politiques publiques.

La donnée ne doit plus être perçue seulement comme un élément élémentaire d'un système d'information, mais bien comme un actif stratégique de nos missions. Le législateur a perçu l'intérêt des données en l'intégrant quasi systématiquement dans les lois récentes ou à venir.

Pour fixer un cadre de travail commun autour de la donnée, **la feuille de route 2020/2021 dresse 25 actions ministérielles concrètes** et organisées en trois dimensions liées entre elles.

### 1. le « capital humain »

La connaissance, les compétences méthodologiques, l'expertise technique ou juridique sont centrales pour comprendre les évolutions des métiers qui sont impactés par la donnée et les innovations numériques, pour identifier et développer des usages, comme la détection de la fraude, la recherche d'entreprises potentiellement en difficulté, l'amélioration du recrutement ou des achats.

Des moments d'**acculturation**, des **ateliers techniques**, des **nouvelles formations dédiées aux traitements des données** ou des **séquences de mentoring adaptées aux managers et l'encadrement supérieur** seront proposées depuis janvier 2020 pour développer une **culture commune de la donnée** au sein de Bercy et **créer une cohérence entre les métiers, les formations et le parcours professionnel**.

### 2. l'« architecture de la donnée »

La donnée impose un cadre structuré avant l'exploration du champ des possibles : une organisation claire, des définitions métiers précises et partagées, une connaissance du patrimoine de données, et en particulier des référentiels, etc. Sans cela, il perdure des risques de doublons, de la complexité inutile, des coûts importants, des incohérences entre les bases ou encore de difficultés de mises à jour.

Il s'agit notamment d'installer **une gouvernance de la donnée**, qui incarne la dynamique de transformation, de **référencer les données** (ie. connaissance des référentiels et des données de références), de **normaliser les définitions et les concepts** utilisés par les données, et enfin de **garantir un usage des données** dans les règles de sécurité et de conformité, comme le bon usage des possibilités offertes par les technologies du *cloud* en fonction de la criticité des données. Ces actions parfois complexes seront mises en place progressivement sur les deux années pour **gouverner, connaître, partager et ouvrir les données**.



### 3. L'« usage des données »

Partie visible de l'iceberg dans l'océan des données, **les usages révèlent la valeur parfois cachée**. La richesse des bases de données réside dans la multiplicité des usages qui peuvent en être faits.

La mise en place d'un **datalake interne** dédié aux données d'entreprises pour améliorer la coordination de nos services, l'animation d'une **communauté de datascientists** pour identifier les meilleurs cas d'usages de la donnée et de l'intelligence artificielle, la **valorisation de bases de données** pour simplifier des processus internes, favoriser les initiatives comme celles des civic-tech ou encore créer de nouveaux services numériques – données essentielles de la commande publique, référentiel des bâtiments, référentiel des ressources humaines, etc. – seront des actions concrètes à mener sur les deux années.

Les 25 actions seront **suivies au niveau de la communauté des administrateurs des données du MEFR**, tous les trimestres, afin d'évaluer leur avancement, les éventuelles difficultés rencontrées et les impacts tangibles dans la transformation publique de Bercy.



## CONTEXTE

La donnée ne doit plus être perçue seulement comme un élément du système d'information, mais bien comme un actif stratégique et essentiel du fonctionnement de l'économie et de l'État. Les lois récentes ou à venir – loi pour une République numérique, loi relative à la transition énergétique pour la croissance verte (TEPCV), loi d'orientation de la mobilité, loi de finances 2020 sur la détection de la fraude fiscale grâce aux informations publiées sur les réseaux sociaux... – intègrent un volet « données » dans la conduite des politiques publiques et dans le rôle régulateur de l'État.

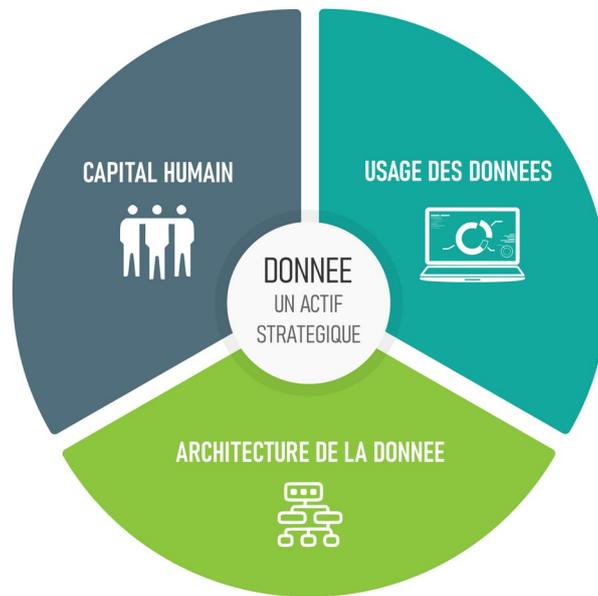
En mars 2019, les ministères économiques et financiers ont créé la fonction d'administrateur ministériel des données (AMD) pour mobiliser tout le potentiel des données au service des politiques publiques, et ainsi donner une nouvelle impulsion dans toutes les composantes des missions de Bercy.

Pour créer cette impulsion, un réseau des administrateurs des données nommés par les directions et services de Bercy a été mis en place. En parallèle, plusieurs actions inédites ont été réalisées au cours de l'année 2019 : des conférences-débats dédiées aux données et aux technologies émergentes (comme l'intelligence artificielle, la datascience ou la blockchain), une journée de la donnée (29 novembre 2019) avec 2 grands témoins, deux tables rondes et près de 30 ateliers, un accompagnement aux appels à projet organisés par la DITP ou la DINUM (AMI IA2, FAST, FTAP, etc.), l'incitation à la création de start-up d'État à dominante données.

Pour 2020-2021, l'AMD propose de lancer des premières actions d'une nouvelle ambition commune de la donnée, que les directions et services devront décliner dans leur programme de travail, ceux-ci étant en première ligne de la production et de l'utilisation des données dans leurs missions. Ainsi, en fonction de leurs priorités, chaque entité pourra intégrer la donnée dans son ADN.

Cette feuille de route dresse 25 actions identifiées au sein du réseau des administrateurs de données ; elles sont organisées en trois dimensions qui ont une propriété transversale à l'ensemble des métiers et des politiques publiques traitées au sein du ministère de l'économie, des finances et de la relance :

- la dimension « capital humain » : avec une montée en connaissance des agents du MEFR pour qu'ils puissent mobiliser les données dans leurs missions ou comprendre comment les données impactent leur travail de tous les jours ;
- la dimension « architecture de la donnée » : avec une gouvernance, qui incarne la dynamique de transformation au plus haut niveau dans toutes les composantes organisationnelles du MEFR, le référencement des données (ie. connaissance des référentiels et des données de références), la normalisation des données et également la sécurité et la conformité pour garantir un usage des données dans les règles ;
- la dimension « usage des données » : avec des méthodes et des outils qui permettent l'émergence d'une offre de services pour maîtriser le cycle de vie de la donnée (collecter, traiter, gérer, diffuser et valoriser les données).



Ces actions sont priorisées pour s'inscrire sur la période 2020-2021 en prenant en compte les attentes des directions et services.

- Les actions qui sont à conduire tout au long de la période, comme celle relevant de l'acculturation, sont notées **(CONT)** pour « action continue »;
- Les actions qui sont à réaliser à un moment de la période sont notées **(20S1)** pour le premier semestre de l'année 2020, **(20S2)** pour le second semestre de l'année 2020, etc.

### *Clefs de succès*

Deux conditions sont particulièrement nécessaires pour relever le défi de l'agenda de la feuille de route ; elles devront être traitées au niveau des enjeux d'organisation qu'elles portent.

La première concerne la mobilisation de ressources, dont d'une part celles des directions qui doivent permettre à leurs agents de participer, contribuer et piloter certaines des actions, et d'autre part, celles du champ technique pour accompagner les expérimentations et les projets à dominante données par nature exigeant en capacité de traitement, en stockage et en disponibilité rapide.

La seconde concerne la capacité des équipes à travailler ensemble, dans un esprit de partage et de bienveillance, en favorisant des développements par produits et par itération, alliant tous les métiers nécessaires techniques ou fonctionnels.

### *Donnée (périmètre de la feuille de route)*

Par donnée, il est entendu une description élémentaire de nature numérique d'une réalité (chose, événement, mesure, observation, transaction, etc.) et représentée sous forme codée. Exemple : une base de données, un indicateur, un texte, une image...



### *Résilience*

Le confinement général activé lors de la crise sanitaire débutée fin 2019 en Chine a permis la mise en avant du télétravail et du numérique (internet, outils collaboratifs, webconférence, etc.). Ainsi le poste de travail numérique des agents de l'État, couplé à des outils comme Tchap, Jitsi, Visiby, a démontré qu'il était un atout dans la continuité des missions indispensables ou essentielles dans une situation de crise inédite.

Cependant, les accès aux données des systèmes d'information restent par principe relativement interdits en mode télétravail tant la crainte d'un accès illégitime est forte. Ainsi, les données accessibles en télétravail sont peu existantes nativement, ce qui a perturbé certaines activités pendant le confinement. Des actions ont été prises dans l'urgence pour assurer les missions du plan de continuité d'activités (PCA).

Dans une logique miroir au plan de reprise d'activité, une lecture orientée donnée est nécessaire pour identifier les fonctions « données » essentielles ou non à l'activité. Cette analyse est à conduire par le prisme de la gouvernance de la donnée, et en particulier sur la connaissance du patrimoine de données (action n°12).

Sans le numérique, la société tout entière serait à l'arrêt. Sans les données, les décideurs seraient aveugles.



## Sommaire

Contexte.....	6
I. Capital humain.....	10
Développer une culture commune de la donnée.....	10
Créer une cohérence entre les métiers, les formations et le parcours professionnel.....	11
II. Architecture de la donnée.....	12
Gouverner la donnée.....	12
Connaître la donnée.....	13
Partager et ouvrir les données.....	16
III. Usages des données.....	18
Proposer des outils de première nécessité.....	18
Favoriser la science de la donnée.....	18
Accompagner les usages innovants des données.....	19
Annexes.....	21
Liste des actions.....	21
Réseau des administrateurs de données.....	22
Liste des abréviations.....	23



## I. CAPITAL HUMAIN

*Acculturer, sensibiliser, susciter, inspirer, impliquer, former*

La connaissance et les compétences techniques, méthodologiques ou juridiques sont centrales pour comprendre les évolutions des métiers impactés par la donnée et les innovations numériques, pour identifier et développer des usages, comme la détection de la fraude, la recherche d'entreprises potentiellement en difficulté, l'amélioration du recrutement ou des achats.

### DÉVELOPPER UNE CULTURE COMMUNE DE LA DONNÉE

- **Acculturer les agents aux données et au champ des possibles** (action 1) (CONT)

La connaissance des possibilités offertes par les données et les innovations de plus en plus matures, comme l'intelligence artificielle, est une priorité pour permettre aux agents de se projeter dans l'évolution de leurs métiers.

En lien avec les directions et services de Bercy, la sphère interministérielle, mais aussi des acteurs privés, plusieurs événements ouverts à tous les agents sont proposés pour les acculturer aux données : des **conférences-débats** mensuelles sur le thème des données et des innovations qu'elles induisent et, une fois par an, la « journée de la donnée » des ministères économiques et financiers.

#### Journée de la donnée 2019

Le SG du MEFR a organisé le 29 novembre 2019 une journée dédiée à la donnée, qui a réuni 300 participants au centre de conférences Pierre Mendès France. Deux grands témoins : Jérôme Nika chercheur en musique et Luc Julia directeur numérique de Samsung. Deux tables rondes : la première sur l'intelligence artificielle au service de nos missions, la seconde sur les compétences à détenir. Trente ateliers pour partager les projets et les initiatives des directions.

Par ailleurs, des **séquences d'information et de sensibilisation** sont également proposées chaque mois pour aider les agents à appréhender les enjeux de la donnée. Ces séquences sont

généralistes (par exemple : sensibilisation aux secrets, au RGPD, à la loi de 1951) et s'adressent à l'ensemble des profils des agents, qu'ils soient managers ou non, spécialistes de la donnée ou non. Ces séquences doivent permettre de placer la donnée dans la chaîne de valeur d'élaboration d'un service public et d'inspirer de nouveaux usages des données au sein d'un service.

- **Organiser des « déjeuners techniques »** (geek-lunchs) et des concours (action 2) (CONT)

Des événements plus spécialisés sont également proposés chaque mois pour découvrir des outils, des technologies utiles, voire émergentes, pour traiter de la donnée. Ces moments de partage d'expertise permettent d'approfondir les connaissances et de détecter les expérimentations à conduire pour les mettre en œuvre dans les métiers.



## CRÉER UNE COHÉRENCE ENTRE LES MÉTIERS, LES FORMATIONS ET LE PARCOURS PROFESSIONNEL

Ces actions ont pour but de proposer un nouveau cadre pour donner de la clarté RH dans un domaine par nature complexe et pour rendre attractive la filière « donnée et numérique » notamment en lien avec les formations initiales et de reconversion (action TA 3). L'objectif est d'identifier les besoins en compétence « données » au sein des directions de Bercy, que les agents soient spécialistes ou non des données, et de construire des parcours de formations adaptés.

- **Définir les compétences et les métiers de la donnée** (action 3) (20S1)

Le métier de *datascientist* est réputé le plus attractif, mais il n'existe pas de définition univoque et partagée de ce métier et des compétences recherchées. D'autres métiers de la donnée – data-ingénieur, data-analyst, data-juriste pour n'en citer que trois – émergent également sur tout le cycle de vie de la donnée et nécessitent également une définition commune au MEFR.

Pour assurer une clarté dans le recrutement de talents, dans la professionnalisation et dans la gestion des compétences, une cartographie des métiers de la donnée et des compétences associées doit être dressée dans le contexte de Bercy. Inspiré des travaux de grandes entreprises du CAC40, du CIGREF ou du référentiel interministériel des métiers de l'État (RIME), ce référentiel des métiers de la donnée à coproduire avec les directions apportera de la cohérence dans la recherche de compétences, qu'elles soient détenues en interne ou en externe au MEFR. Ce référentiel sera diffusé auprès de la DINUM qui assure un rôle de coordination des métiers du numérique avec la DGAFP.

### Place des femmes dans la donnée et l'intelligence artificielle

Pour inspirer toutes les femmes du MEFR dans le domaine de la donnée et de l'IA (et en lien avec le chantier TA3 de la feuille de route Tech.Gouv de la Dinum), le SG s'engage à la parité femme-homme dans tous ses événements (tables rondes, conférences-débats).

- **Enrichir le programme de formation et créer un vivier de compétences internes** (action 4) (CONT)

Selon le Gartner<sup>1</sup>, se former en 2020, c'est améliorer son employabilité pour les cinq prochaines années maximum au lieu de trente-cinq au début du 20<sup>e</sup> siècle. Les métiers évoluent donc très vite et il est nécessaire de s'y préparer.

En lien avec l'IGPDE, il s'agit de revoir chaque année les formations proposées et de les organiser selon un parcours adapté aux métiers du MEFR qu'ils soient spécialistes ou non de la donnée. De nombreux métiers sont d'ores et déjà bousculés par les données ; former les agents publics managers ou non le plus souvent possible et en amont de leurs prochaines missions permet de mieux les préparer aux futurs métiers actuellement en tension, et de les mettre en condition favorable pour améliorer leur employabilité et répondre à leurs besoins d'évolutions professionnelles.

Une expérimentation est également à proposer pour certifier des agents aux différents métiers de la donnée.

Cela permet de donner plus de cohérence dans l'acquisition de connaissances des agents, des managers et de les former selon une approche collective et moins individuelle, comme cela peut être encore pratiqué.

- **Développer et promouvoir une offre de formations en ligne** (action 5) (CONT)

Dans le domaine de la donnée et en particulier des langages R et Python plébiscités par la communauté des datascientists, l'offre de formation en ligne est particulièrement importante : des

1 Source : 2019 Gartner CEO and Senior Business Executive Survey



plateformes dédiées, comme [DataScientest.com](https://www.datascientest.com), émergent, des plateformes généralistes proposent des cursus autour des données et de la datascience, et sans oublier le ministère en charge de l'écologie qui a publié une [valise pédagogique](#) sur le langage R apprécié des experts de la donnée.

L'offre de formation doit ainsi intégrer les principaux outils de la donnée, y compris les plus émergents pour permettre une montée en connaissance rapide des agents.

Des sensibilisations aux différents secrets et au RGPD seront également à proposer pour développer une meilleure connaissance des principales règles qui régissent la protection des données.

En collaboration avec l'IGPDE, il s'agit d'identifier les principales offres, de les expérimenter auprès d'experts et d'en assurer la diffusion à un plus large public. La création d'un « passeport de la donnée » est à étudier. Il permettrait d'acter la reconnaissance de l'administration dans le parcours de formation des agents et ainsi valoriser les nouvelles connaissances acquises.

L'offre de formations doit être adaptée par défaut au télétravail.

## II. ARCHITECTURE DE LA DONNÉE

*Organiser, structurer, référencer, partager, ouvrir, qualifier, qualité*

### GOUVERNER LA DONNÉE

Comme tout actif vital d'une entreprise ou d'une administration, le pilotage de la donnée doit être assuré par une gouvernance, qui identifie, classe, pilote et contrôle les données en s'appuyant sur des processus, des règles, des acteurs et des instances de décision. Cette gouvernance priorise les actions, en particulier sur les données à forte valeur, et identifie les moyens pour produire, mettre en qualité ou mettre à disposition les données.

- **Intégrer la donnée dans la gouvernance stratégique** de Bercy (action 6) **(20S1)**

La maîtrise de la donnée est un des enjeux du ministère de l'économie, des finances et de la relance. Impulsée par le décret<sup>2</sup> de création de la Direction interministérielle du numérique, elle s'intègre pleinement dans la mutation des DSI ministérielles en direction du numérique<sup>3</sup> et elle se décline dans l'ensemble de la comitologie stratégique et opérationnelle.

Le SG veille à intégrer la donnée dans la gouvernance ministérielle, à l'instar des politiques transversales comme les ressources humaines, et en priorité dans les nouveaux chantiers comme « Bercy Vert ».

Par ailleurs, un accompagnement spécifique est proposé aux directeurs d'administration centrale pour les aider à intégrer la donnée comme un nouveau moyen pour piloter les politiques publiques, dont ils ont la responsabilité. Une revue trimestrielle des actions de la feuille de route est présentée auprès des directeurs d'administration centrale.

2 <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000039281619>

3 Arrêté du 18 décembre 2019 portant organisation du secrétariat général des ministères économiques et financiers <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000039675844&categorieLien=id>



- **Animer une communauté des administrateurs des données** par direction et service (action 7) (CONT)

La richesse des métiers et des missions ne peut se traiter par un unique administrateur pour Bercy ; c'est un travail collégial qui doit inclure chaque direction et chaque service du MEFR avec leurs priorités et leur sensibilité aux données. Les services ont été invités à nommer courant été 2019 un administrateur des données. Ce réseau permet de fiabiliser les travaux et de construire un dialogue permanent au sein du ministère de l'économie, des finances et de la relance, mais également au travers de l'AMD avec les acteurs de la DINUM et les autres ministères.

Animée par le SG, cette communauté se réunit trois à quatre fois par an et assure le pilotage de la feuille de route « donnée » de Bercy.

D'autres réseaux – correspondants opendata, correspondants innovations numériques – sont également animés par le SG ; ils se réunissent en fonction des sujets à traiter (au minimum une fois par trimestre).

- **Développer l'échange de données par API** (action 8) (CONT)

Le recours aux API favorise l'échange, la mise à disposition de données de qualité et l'exploitation des données. Les systèmes d'information historique doivent s'API-ser pour converger vers la logique d'État-plateforme et proposer des API pour exposer leurs données ou accéder à des données externes.

Pour organiser les échanges de données au sein du MEFR, mais aussi avec nos partenaires externes (ministères ou ayant droit externes), un comité de pilotage API permet d'assurer une coordination de ce chantier dans un esprit de simplification de l'exposition des API en prenant en compte les spécificités des données. Il produira une doctrine en la matière. Ce comité se réunit tous les trimestres.

Pour le volet interministériel, en lien avec le chantier DA2 de la feuille de route Tech.Gouv, le SG soutient la solution de management d'API PISTE, opérée par l'AIFE, car avec près de 15 millions d'appels par jour, cette solution propose un niveau de service industriel et elle est retenue par plusieurs de nos partenaires pour publier leurs API. Au titre du FTM 2020, le SG finance la publication d'API sur PISTE.

Par ailleurs, un forum « API-culture » sera organisé chaque année en juin pour encourager les fournisseurs de données à créer des API et les utilisateurs de service à les utiliser.

- **Poser un cadre d'usage du *cloud* pour les données** (action 9) (20S2)

Piloté par la Dinum, le *cloud* de l'État se développe et son adoption par les différentes directions et services du MEFR nécessite des réflexions pour prendre de manière cohérente les bonnes décisions pour l'hébergement des données du MEFR. Le modèle hybride (mix entre le cloud en propre et le cloud externalisé) est un des modèles préconisés pour répondre aux exigences de sécurité des différentes données, dont le MEFR est amené à traiter. Or en première analyse, ce modèle peut apparaître le moins pertinent financièrement, car il somme les coûts d'investissement sur deux hébergements distincts. Il est donc nécessaire de forger un avis éclairé par plusieurs expérimentations des directions au regard de la sensibilité des données, des traitements qu'elles nécessitent, des doctrines actuelles d'hébergement et enfin des outils en cours de montée en charge, comme :

- l'offre de services cloud « NUBO » de la DGFIP
- la plateforme « DDODS » de l'Insee



## CONNAÎTRE LA DONNÉE

Bien avant l'utilisation des données, il est nécessaire de développer une vision la plus exhaustive possible et partagée des données détenues au sein de Bercy. Dans le rapport Fouilleron, la multiplicité des bases de données pour un même objet métier est clairement établie, créant alors des effets pervers, comme l'augmentation des coûts, l'incohérence entre les bases ainsi créées, la difficulté de les mettre à jour, leur caractère de référence, etc. Le MEFR doit lutter contre cette prolifération incontrôlée des données, et en premier lieu en partageant les référentiels.

L'architecture de la donnée nécessite une cartographie des données, librement accessible et à jour, qui permet le classement, l'organisation et l'écoulement des données selon des critères partagés et avec la prise en compte de la dimension métier portées par les données. C'est également un chantier relativement chronophage qui nécessitera une mise en œuvre pragmatique, démarrant par les référentiels.

- **Développer le service public de la donnée** (action 10) (20S2)

Créé en 2016 par la loi pour une République numérique avec neuf premières sources, le service public de la donnée (SPD) vise à mettre à disposition des jeux de données de référence qui présentent un enjeu le plus fort. Or, aucune autre donnée n'a été pour le moment identifiée. Pourtant, il existe d'autres données de référence, comme les bâtis, les données essentielles de la commande publique, les consommations énergétiques de bâtiments de l'État, l'identité numérique des agents de l'État.

Un travail spécifique d'accompagnement est également nécessaire pour identifier de nouvelles données d'intérêt à intégrer dans le SPD, ainsi que le recours du SPD dans les systèmes d'information existants ou à venir.

### Service public de la donnée : des données sur lesquelles vous pouvez compter

Le service public de la donnée créé par l'Article 14 de la loi pour une République numérique vise à mettre à disposition, en vue de faciliter leur réutilisation, les jeux de données de référence qui présentent le plus fort impact économique et social. Il s'adresse principalement aux entreprises et aux administrations pour qui la disponibilité d'une donnée de qualité est critique. Les producteurs et les diffuseurs prennent des engagements auprès de ces utilisateurs. La mission Etalab est chargée de la mise en œuvre et de la gouvernance de ce nouveau service public. Elle référence l'ensemble des données concernées sur cette page.

**Les données de référence**

A ce jour, neuf jeux de données, qui couvrent un large champ thématique ont été identifiés comme des données de référence.

<p><b>Base Adresse Nationale (BAN)</b> La Base Adresse Nationale est une base de données qui a pour but de référencer l'intégralité des adresses du territoire français. Elle contient</p> <p>1 33 40 France</p>	<p><b>Base Sirene des entreprises et de leurs établissements (SIREN, SIRET)</b> Pour vous abonner à notre lettre d'information Sirene open data actualités, suivez ce lien : <a href="https://insee.fr/fr/information/1405555">https://insee.fr/fr/information/1405555</a> Pour</p> <p>15 7 27 France</p>	<p><b>Code Officiel Géographique (COG)</b> Chaque année, l'Insee met à disposition sur son site (Insee.fr) le code officiel géographique qui rassemble les codes et libellés des communes.</p> <p>35 12 22 France @ Commune français</p>
<p><b>Plan Cadastral Informatisé (PCI)</b> Plan Cadastral Informatisé PCI Ce jeu de données provient d'un service public certifié Données de référence Le plan cadastral est un</p> <p>3 57 France</p>	<p><b>Registre parcellaire graphique (RPG) : contours des parcelles et îlots cultureux et leur groupe de cultures majoritaire</b> Le registre parcellaire graphique est une base de données géographiques servant de référence à l'instruction des aides de la politique agricole</p> <p>20 2 12 2013-2017 France</p>	<p><b>Référentiel de l'organisation administrative de l'Etat</b> Le décret du 14 mars 2017 a institué le Service Public de la Donnée. Celui-ci met à la disposition du public 9 jeux de données de</p> <p>5 4 16 2010-2025 France</p>
<p><b>Référentiel à grande échelle (RGE)</b> L'État a confié à l'IGN le développement du référentiel à grande échelle (RGE). Pour ce faire, il fait appel à ses moyens propres ainsi</p> <p>12 2 6 France</p>	<p><b>Répertoire National des Associations (RNA)</b> Que contient le RNA ? Le RNA répertorie l'ensemble des associations relevant de la loi du 1er juillet 1901 relative au contrat d'association.</p> <p>19 5 23 1901-2017 France @ Point</p>	<p><b>Répertoire Opérationnel des Métiers et des Emplois (ROME)</b> Dans un contexte marqué par de fortes mutations de l'environnement économique et social, le ROME (Répertoire Opérationnel des Métiers et</p> <p>5 2 3</p>

La coordination autour de standards et de directives européennes (par exemple la directive « public sector information » mise à jour en juin 2019) est à animer pour garantir l'accès libre et gratuit aux données « de grande valeur ».



- **Définir une norme de la donnée** (action 11) (20S2)

La normalisation et les standards précisent des règles d'écriture des données afin de faciliter les échanges entre les systèmes d'information<sup>4</sup>. Cette démarche favorise l'interopérabilité et la création de nouveaux services pour les agents du MEFR, mais aussi pour les usagers.

En suivant les mêmes principes de réalité que ceux préconisés pour l'inventaire des données, un travail de fond est à engager en collaboration avec l'ensemble des directions, en commençant par identifier les normes existantes (notamment les normes archivistiques) et celles qui pourront être mutualisées pour profiter à tous.

- **Référencer de manière pragmatique les données détenues**, en priorité les référentiels, par les directions et les services (action 12) (20S2)

Un inventaire des données détenues par les services de Bercy est un chantier ambitieux, complexe et continu. Une approche non dogmatique est nécessaire pour construire une connaissance des données sans épuiser les ressources. L'utilisation d'un standard de catalogage (DCAT par exemple) sera privilégiée pour faciliter la découvrabilité et le moissonnage des données cataloguées.

Bien que non exhaustives, plusieurs pistes sont à explorer : se focaliser dans un premier temps sur les données de type référentiel, puis l'enrichir grâce à l'inventaire des systèmes d'information ou des données ouvertes de la [plateforme « data.economie »](#).

Cette première version de l'inventaire doit permettre la classification de ces données, notamment sur la sensibilité des données, pour mieux éclairer les agents à leur usage, y compris en prenant connaissance de certaines limites portées par le RGPD ou les différents secrets, comme celui instauré par la loi du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques.

- **Urbaniser des données** (action 13) (21S1)

Depuis 2012, sous l'impulsion de la DISIC devenue Dinum, les ministères économiques et financiers ont mis en place un plan d'occupation des sols pour organiser le système d'information de l'État, véritable support technique pour stocker et utiliser les données.

Cette vision urbanisée du patrimoine de données doit permettre de mettre en cohérence l'organisation des données pour éviter les doublons et faire émerger des données de références, qui *in fine* ont vocation pour les plus partagées d'entre elles à intégrer le [service public de la donnée](#) (cf. action 10).

### Utilisation du répertoire Sirene dans des applications

Les équipes de SEP1 mettent en œuvre le répertoire Sirene dans plusieurs de ses applications, comme Cassiope, Egide et Geida. Cela permet d'assurer une cohérence de la donnée des entreprises et facilite l'interopérabilité entre les applications. D'autres directions, comme la DGFIP ou la DGDDI, utilisent ce référentiel pour assurer la même cohérence.

*Cassiope : Application de gestion des zones à régime restrictif (ZRR) pour le SHFDS*

*Egide : Application de dématérialisation de l'instruction des demandes de licences pour l'exportation de biens à double usage*

*Geida : Application de dématérialisation des procédures devant les commissions d'aménagement commercial (CIMAP)*

### La terminologie

Le ministère de l'économie, des finances et de la relance a élaboré un référentiel terminologique sur le vocabulaire en général, et celui de la donnée en particulier. Il permet la bonne compréhension des termes pour tous. Il est aujourd'hui accessible sur le site terminologie et sera prochainement enrichi de nouveaux termes, comme *datalake*, *dataops*, *dataviz*, API, pile IA, *tracking*, *tracing*, etc.

4 Cf. guide ISO/CEI 2 : <https://www.iso.org/fr/standard/39976.html>



Il est attendu que chaque direction se positionne sur un jeu de données en cohérence avec leurs programmes de travail de 2020 et de 2021.

### Identité numérique des agents, entreprises et particuliers

Pour améliorer la sécurité des accès aux données et aux systèmes d'information, le SHFDS propose la mise en place de la carte « rossignol » à l'instar d'autres directions ou d'autres ministères. Cette question de l'identité numérique professionnelle et de la maîtrise des modèles des droits applicatifs d'accès aux applicatifs est certainement un des sujets de données est des modèles les plus communs dans tous les SI et englobé dans l'approche France Connect et qui posent des défis de maîtrise de référentiels aux projets et aux administrations des données.

### Données essentielles de la commande publique

Depuis le 1<sup>er</sup> octobre 2018, les acheteurs publics sont tenus de publier sur leur profil d'acheteur les données essentielles de marchés publics et des contrats de concession qu'ils attribuent. À moyen terme, les données de recensement seront également proposées. Cette exigence des données, double numérique de la commande publique, participe à sa transparence. Néanmoins, les données sont aujourd'hui produites par différents canaux, ce qui complexifie l'objectif d'une donnée consolidée, exhaustive et de qualité (sans doublon notamment). En associant les éditeurs et en définissant une stratégie de la donnée, ce projet permet d'assurer une mise en cohérence des données en lien avec les nouveaux usages qu'elle permet par son ouverture.

## PARTAGER ET OUVRIR LES DONNÉES

- **Partager le processus du cycle de vie de la donnée** (action 14) (20S1)

Le processus du cycle de vie de la donnée<sup>5</sup> couvre un champ assez large : la collecte, le stockage, le traitement, la valorisation et l'archivage des données. Ce processus doit pouvoir être défini dans les grandes lignes pour tous les services de Bercy, et tout particulièrement les directions comme l'Insee dont les équipes traitent un très grand nombre de données et ont une connaissance très précise de leurs partenaires.

- **Mettre en qualité des données** (action 15) (CONT)

Les données sont la base de nombreuses analyses et à l'origine d'autant de décisions. Cependant, des données imprécises, non exhaustives ou simplement erronées conduisent à des résultats inexploitable, voire dangereux. Par conséquent, il est nécessaire de s'appuyer sur une qualité des données avant d'en extraire des informations exploitables. En s'appuyant sur les travaux menés dans le cadre de la publication des données ouvertes, et en collaboration avec les producteurs de données du MEFR, un groupe de travail sera organisé avec les directions pour définir un processus de gestion de la qualité elle-même adaptée aux usages attendus des données.

Chaque direction et service du MEFR s'engage à documenter un à deux jeux de données par an.

### Les listes électorales

Jusqu'en 2018, les listes électorales gérées par chacune des 35 500 communes françaises contenaient de nombreuses anomalies relevées régulièrement par la presse ou les citoyens : inscriptions en double, décédés figurants sur les listes... Depuis mars 2019, les listes électorales sont extraites du Répertoire Électoral Unique tenu par l'Insee et garantissant notamment l'unicité d'inscription, grâce à un travail inédit de mise en qualité des données.

5 Gestion du flux de données d'un système d'information tout au long de son cycle de vie qui permet de trier les données en fonction de leur ancienneté, dans le but d'améliorer les performances, de réduire les coûts de stockage, et de se conformer aux lois et réglementations sur la gouvernance des données.



- **Créer une expertise pour anonymiser les données** (action 16) (20S2)

L'ouverture des données impose sauf avis contraire du législateur (exemple : les données de décès ou les données des permis de construire) la protection des données personnelles et par construction l'impossibilité de pouvoir retrouver un individu en particulier. Pour y parvenir, les données doivent faire l'objet d'un traitement d'anonymisation par le producteur. En concertation avec le département de la méthodologie de l'Insee (voire également par le centre d'expertise de la statistique européenne<sup>6</sup>), un groupe de travail est à mettre en place pour partager les bonnes pratiques et apporter conseils et exemples.

- **Mettre en place une chambre d'enregistrement des demandes de partage de données** (action 17) (20S2)

Pour faciliter le partage des données entre directions du MEFR, une chambre d'enregistrement et de suivi des demandes doit être mis en place. En particulier, elle assurera une mutualisation des questions juridiques au regard du droit de la donnée émergent et parfois complexe. Des premières demandes ont d'ores et déjà été identifiées par des directions générales, comme celle de la CCRF pour des besoins de profilage sur la question de la fraude.

- **Animer une communauté des données ouvertes** (action 18) (20S1)

Malgré plus de 365 données ouvertes par les directions et services du MEFR sur le portail [data.economie.gouv.fr](https://data.economie.gouv.fr), il reste difficile d'être à l'écoute des utilisateurs de ces données, faute de communauté organisée et représentative des usages possibles. Pour y remédier, le SG du MEFR initie une communauté qui rapproche les producteurs et utilisateurs de données.

Un rendez-vous annuel, de type « non conférence » est proposé en fin d'année et donne la parole aux membres de la communauté. Les travaux qui y seront réalisés seront proposés par les participants et non listés à l'avance par les directions du MEFR.

### III. USAGES DES DONNÉES

*Outils, valoriser, mêler*

#### PROPOSER DES OUTILS DE PREMIÈRE NÉCESSITÉ

- **Expérimenter la mise en place d'un datalake interne dédié aux données d'entreprises** (action 19) (21S2)

Aujourd'hui, pour des raisons historiques, les données sont stockées dans des silos de données étanches entre eux et très souvent avec des outils de Business Intelligence traditionnels qui limitent l'exploration des données. La construction d'un datalake s'impose comme un outil incontournable d'une stratégie de la donnée.

---

<sup>6</sup> Des outils de gestion du secret statistique sont mis à disposition sur internet : <https://joinup.ec.europa.eu/solution/sdctoolstools-statistical-disclosure-control>



Pour les missions de pilotage budgétaire par exemple, le principal enjeu est de disposer d'une vision à « 360° » des dépenses, des délais de paiement, de l'évolution de la consommation par rapport à l'année dernière, avec une expérience utilisateur personnalisée et avec des données collectées en temps réel. Le datalake répond à un besoin de « réservoir » de données favorisant analyses et modèles de décision pour les budgétaires, également pour la gestion des ressources humaines ou le contrôle de gestion, et plus largement pour de nombreux métiers du MEFR.

Au sein du MEFR, la quasi-totalité des directions et services créent ou utilisent des données sur les entreprises, comme le référentiel Sirene géré par l'Insee et aujourd'hui accessible par API. Des directions volontaires seront associées pour créer un espace de partage d'informations structurées autour des entreprises au sein d'une communauté connue et légitime d'en connaître pour leurs missions. Au-delà des aspects techniques qu'il faut résoudre, en privilégiant l'open source via un partenariat avec d'autres acteurs comme le [TOSIT](#), le principe objectif est de construire une communauté avec des règles et des conditions d'accès aux données des entreprises entre partenaires internes en cohérence avec les éventuels secrets liés aux données (RGPD, CNIL).

- **Prototyper une architecture « data-ops »** (action 20) **(21S1)**

Le « data-ops » est une méthodologie collaborative de gestion des données, dont l'objectif est d'améliorer la communication, l'intégration et l'automatisation des flux de données entre les gestionnaires et consommateurs de données au sein d'une organisation. (*source : Gartner*)

Plusieurs défis sont à relever : l'association des différents métiers et des équipes, l'efficacité du processus et la diversité des technologies à mettre en œuvre.

Une expérimentation doit permettre de tester une première version d'une architecture « data-ops », avec l'apport méthodologique de l'Insee, de la Douane et de la DGFiP. Des outils comme la distanciation (avec les outils de l'IGN par exemple) ou la data-visualisation (comme le projet DATA360<sup>7</sup>) sont à proposer pour créer des automatismes sur le cycle de vie de la donnée.

## FAVORISER LA SCIENCE DE LA DONNÉE

- **Animer une communauté de la datascience** (action 21) **(CONT)**

Depuis près d'un an, les directions et services recrutent de profils spécialisés dans le traitement des données (data-analyste, data-ingénieur, data-scientists), mais ils restent isolés dans leur structure. Le besoin de communiquer entre eux est fort et la création d'une communauté doit permettre de favoriser l'inspiration croisée des équipes et encourager les directions non pourvues de ces profils à recruter.

Cette communauté sera réunie tous les mois à Bercy et ce moment sera dédié aux partages des projets en cours et aux outils utilisés. Il s'agit essentiellement d'un moment d'échange informel et de réseautage.

- **Expérimenter des innovations autour de l'IA** (action 22) **(20S1)**

Plusieurs besoins ont été évoqués par les directions : la détection des entités nommées (quasi utile à toutes les directions), la détection des clauses abusives dans les contrats (CCRF), la détection de fraudes depuis les réseaux sociaux, le rapprochement de données sans clef d'appariement. De nombreuses solutions sont en cours de développement chez des éditeurs et seront matures dans les prochains mois. En même temps, des développements sont également réalisés par des équipes multi-profils.

---

7 L'objectif du projet DATA360 est de proposer aux décideurs, chefs de services et sous-directeurs, des tableaux de bord dynamiques et exploratoires des données détenues par le secrétariat général (comme les données budgétaires, d'achats ou des ressources humaines). Une des forces de cet futur outil est de donner confiance aux décideurs dans les indicateurs proposés en permettant l'accès aux données qui ont été mobilisées pour les produire.



Toutes les directions n'ont pas les mêmes moyens pour y parvenir. Il est nécessaire de rassembler les forces<sup>8</sup> au sein du MEFR, c'est à dire organiser des études exploratoires ou des expérimentations sur l'intelligence artificielle pour construire une connaissance des outils et méthodes mobilisables en fonction des moyens humains pour faire ou des moyens budgétaires pour acheter.

- **Communiquer sur un catalogue d'outils** mis en œuvre au sein du MEFR (action 23) (20S2)

L'offre en outils innovants, opensource ou non, pour traiter la donnée sur l'ensemble de son cycle de vie est particulièrement riche : Tableau, Dataiku, Python, R, Elasticsearch, etc. sans oublier des outils plus classiques, mais encore très utilisés, comme Microsoft Excel et Access, ou encore LibreOffice Calc.

Un catalogue doit permettre de les lister, d'en préciser les cas d'usage au sein du MEFR. Il permettra aux agents du ministère de l'économie, des finances et de la relance de disposer d'un socle technique permettant la standardisation des solutions techniques, langages et des compétences nécessaires pour les maîtriser, tout en privilégiant les outils ouverts.

#### Le cloud de la DGFIP

Deux directions élaborent un catalogue de services centrés pour leurs usagers internes, avec une possibilité d'ouverture en interdirectionnel, voire en interministériel.

La DGFIP a construit un service IaaS « cloud cercle 1 » NUBO en cours de déploiement ; il intégrera en 2020 des services, notamment des piles IA, répondant aux normes de sécurité industrielle de l'État définies par l'ANSSI.

L'Insee développe une plateforme « DDODS » destinée aux utilisateurs du système statistique public en interministériel afin de leur permettre d'expérimenter de nouvelles technologies dédiées à la donnée. Elle sera opérée en « best effort » et fonctionnera de manière déconnectée du SI industriel de l'institut.

## ACCOMPAGNER LES USAGES INNOVANTS DES DONNÉES

- **Mettre en place une expertise datascience** (action 24) (21S1)

La réussite des projets, qui nécessitent de la datascience, sont à accompagner avec volontarisme et coordination entre les directions grâce à un apport en expertise. Le SG propose une ressource partagée dans le domaine de la science des données et de l'intelligence artificielle pour accompagner les directions dans leurs projets. Cette ressource pourra apporter son expertise aux directions qui en feront la demande sur un projet en cours ou à venir.

- **Expérimenter l'incubateur de startup d'État du MEFR**, membre de la communauté beta.gouv.fr (action 25) (20S1)

La coordination et la mutualisation des moyens de transformation numériques doivent se renforcer pour favoriser les réutilisations, valoriser les réussites, apprendre des échecs et ainsi savoir capitaliser au sein d'une organisation bien plus apprenante. De plus les évolutions technologiques qui génèrent ces opportunités de transformation numérique (que ce soit le *cloud*, le *bigdata*, l'IA, le devops, l'IOT, etc.) peuvent être complexes à maîtriser et nécessitent des courbes d'apprentissage importantes, où le cumul des expériences et de l'apprentissage par essai et erreur est essentiel.

Dans les principales missions du MEFR – détection de la fraude, protection des intérêts économiques, connaissance des entreprises, gestion des comptes publics, mais aussi ressources

<sup>8</sup> En lien avec la stratégie nationale en intelligence artificielle <https://www.aiforhumanity.fr/>



humaines, gestion des bâtiments, achats – les défis restent globalement à relever tant l’usage des potentialités du numérique s’étend et constituent autant d’opportunités à saisir et organiser à une échelle appropriée pour qu’ils soient transformateurs.

Par nature, les projets innovants peuvent être incertains, ne rapportent rien ou peu à leur début. Pour limiter les coûts et financer efficacement les projets, il est nécessaire d’accélérer la création de valeur, en remettant en cause les projets tous les semestres et savoir pivoter les objectifs de manière appropriée, voire même stopper les projets rapidement lorsque les attentes ne peuvent être satisfaites.

Créer un incubateur permet la mise en place d’un cadre commun en acceptant cette prise de risques, et ce mode de gestion très particulier, en incluant les utilisateurs et les métiers et en s’appuyant sur les acquis et dynamiques ministériels et interministériels existants (beta.gouv.fr, Tech.Gouv et administration ministériel des données). L’incubateur est le lieu d’échange et d’élaboration d’outils numériques puissamment transformateurs (notamment IA) pour les métiers, construits progressivement au gré des différents projets incubés et rassemblés dans un catalogue (type « IBM Watson »). Une communauté IA intégrée à l’incubateur serait un cadre commun de réflexion pour identifier les cas d’usages pertinents pour le MEFR.

C’est également le lieu d’élaboration de démarches de passage à l’échelle (industrialisation) pour les dossiers complexes. Il regroupe des ressources rares, communes et partagées : coach pilotage, coach technique, *designer thinking*, *UX designer*, architecte, expert sécurité, spécialiste de la donnée, etc.

Ainsi, l’incubateur représente une ressource socle indispensable pour accélérer la transformation numérique du MEFR s’appuyant sur les outils numériques et sécurisant les directions dans leurs initiatives de start-up d’État, qui complétera opérationnellement l’incitation budgétaire que constitue le FTM.

Compte tenu de la structure du MEFR, le mode start-up restera à l’initiative des directions, l’incubateur fournissant cadre et support, le tout inspiré de l’incubateur national beta.gouv.fr et encouragé via le FTM qui finance cet outil commun.



## ANNEXES

### LISTE DES ACTIONS

Les actions continues

- **Acculturer les agents aux données et au champ des possibles grâce aux innovations émergentes** (action 1)
- **Organiser des « déjeuners techniques »** (geek-lunchs) **et des concours** (action 2)
- **Enrichir le programme de formation et créer un vivier de compétences internes** (action 4)
- **Développer et promouvoir une offre de formations en ligne** (action 5)
- **Animer une communauté des administrateurs des données** par direction et service (action 7)
- **Développer l'échange de données par API** (action 8)
- **Mettre en qualité des données** (action 15)
- **Animer une communauté de la datascience** (action 21)

Les actions du premier semestre 2020

- **Définir les compétences et les métiers de la donnée** (action 3)
- **Intégrer la donnée dans la gouvernance stratégique** de Bercy (action 6)
- **Partager le processus du cycle de vie de la donnée** (action 14)
- **Animer une communauté des données ouvertes** (action 18)
- **Expérimenter des innovations autour de l'IA** (action 22)
- **Expérimenter l'incubateur de startup d'État du MEFR**, membre du réseau Beta.gouv (action 25)

Les actions du second semestre 2020

- **Référencer de manière pragmatique les données détenues**, en priorité les référentiels, par les directions et les services (action 12)
- **Développer le service public de la donnée** (action 10)
- **Créer une expertise pour anonymiser les données** (action 16)
- **Mettre en place une chambre d'enregistrement des demandes de partage de données** (action 17)
- **Communiquer sur un catalogue d'outils mis en œuvre au sein du MEFR** (action 23)
- **Poser un cadre d'usage du *cloud* pour les données** (action 9)

Les actions du premier semestre 2021

- **Définir une norme de la donnée** (cf. schéma) (action 11)
- **Urbaniser des données** (action 13)
- **Prototyper une architecture « data-ops »** (action 20)
- **Mettre en place une expertise datascience** (action 24)

Les actions du second semestre 2021

- **Expérimenter la mise en place d'un datalake interne dédié aux données d'entreprises** (action 19)



## RÉSEAU DES ADMINISTRATEURS DE DONNÉES

- **DGTRESOR** : Antonin Aviat, Yannis Kemel
- **DGE** : Raphaël Keller
- **DG INSEE** : Thierry Aouizerate, Arnaud Degorre
- **DGCCRF** : Jérémie Vallet
- **CGEJET** :-
- **APE** : Franck Boiteau
- **SG** : Stéphane Trainel (administrateur ministériel des données)
- **DAJ** :-
- **IGF** :-
- **CGEF** : Christian Avazeri, Luc Degardin
- **CBCM** : Philippe Sauvage
- **DGFIP (yc DIE et DLF)** : Lionel Ploquin
- **DB**: Vincent Amilhaud
- **DGAFF** :-
- **DGDDI** : Christophe de Saint-Viance
- **DNLF** : Johny Marcel
- **TRACFIN** : Delphine Le
- **AIFE** : Laurent Robillard
- **CISIRH** : Philippe Cuccuru



## LISTE DES ABRÉVIATIONS

ADD : Administrateur directionnel des données

AGD : Administrateur général des données

AIFE : Agence pour l'informatique financière de l'État

AMD : Administrateur ministériel des données

AMI : Appel à manifestation d'intérêt

Anonymisation : rendre très difficile voire impossible la « ré-identification » des personnes physiques ou morales

ANSSI : Agence nationale de la sécurité des systèmes d'information

API : *Application programming interface*. Permet l'échange d'informations entre deux logiciels de manière simple et autonome

Bigdata : données en masse ou mégadonnées. Se dit à la fois des données avec certaines caractéristiques (volumétrie, hétérogénéité) et des usages qui peuvent en être fait

CNIL : Commission nationale informatique et liberté

DAJ : Direction des affaires juridiques

DECP : Données essentielles de la commande publique

DGDDI : Direction générale des douanes et des droits indirects

DGE : Direction générale des entreprises

DGFIP : Direction générale des finances publiques

DINUM : Direction interministérielle du numérique

DITP : Direction interministérielle de la transformation publiques

DLNUF : Dites-le-nous une fois

DPD : Délégué à la protection des données

EIG : Entrepreneur d'intérêt général

FAST : Fonds d'accélération des startup d'État

FTAP : Fonds de transformation de l'action publique

FTM : Fonds de transformation ministérielle

IA : Intelligence artificielle

IAAS : Infrastructure As A Service

Insee : Institut national de la statistique et des études économiques

IOT : internet des objets

MEFR : Ministère de l'économie, des finances et de la relance

PAAS : Plateform As A Service

POC : *Proof of concept*, prototype

RGPD : règlement européen pour la protection des données

RGI : Référentiel général d'interopérabilité

RGS : Référentiel général de sécurité

SE : Startup d'Etat

SG : Secrétariat général

SHFDS : service du haut fonctionnaire de défense et sécurité

TEPCV : Territoire à énergie positive pour la croissance verte

TG : Programme Tech.Gouv

TOSIT (The Open Source I Trust) : Association Loi de 1901 de soutien à l'opensource



**MINISTÈRE  
DE L'ÉCONOMIE,  
DES FINANCES  
ET DE LA RELANCE**

*Liberté  
Égalité  
Fraternité*

**Secrétariat  
général**