



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

*Liberté  
Égalité  
Fraternité*

# Faire de la donnée un actif stratégique de l'économie, des finances et de la relance

Feuille de route 2020-2022

Mise à jour : juillet 2021



Faire de la donnée un actif stratégique de l'économie, des finances  
et de la relance

## 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 tôt pour tirer un bilan objectif, mais ce bouleversement inédit révèle comme priorité majeure de la Nation les politiques publiques du ministère : soutenir les entreprises dans leur développement économique ou lors de leurs difficultés, faciliter les échanges de marchandises avec les pays étrangers, produire les indicateurs statistiques et économiques pour aider à la prise de décision ou encore détecter les fraudes pour plus d'équité. Or, pour assurer ces missions avec toujours plus de qualité, de rapidité, d'efficacité, pour développer un service public performant, l'information doit être disponible, partagée, ouverte, les méthodes connues et expliquées.

L'information est donc au centre des défis à relever pour mener à bien nos missions. Les données détenues par les directions et services de Bercy sont nombreuses, de qualité, et sont également très utiles pour les autres ministères et le grand public. Une plus forte utilisation de ces données devient aujourd'hui centrale pour un État plus efficient et un service public plus performant. 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.

Au sein du ministère, la mission d'administrateur ministériel des données (AMD) a été créée en avril 2019 pour **soutenir résolument toutes nos directions pendant cette « révolution de la donnée »**, en permettant une meilleure mobilisation des données dans le pilotage de nos politiques publiques ou dans la création de nouveaux services numériques. Construite initialement dans un esprit de « startup d'État », avec une équipe réduite et immédiatement opérationnelle, l'administration des données devait relever un défi majeur : créer les conditions favorables pour que le MEFR tende vers une meilleure maîtrise de ses données, les utiliser, les partager – y compris dans le respect des secrets légaux et de la protection des données personnelles – pour concevoir et piloter les politiques publiques.

**Depuis deux ans, les actions réalisées, dont l'installation d'une communauté interne des administrateurs directionnels des données, les nombreux événements – deux journées de la donnée, deux forums API-CULTURE et de nombreuses conférences-débats – et des projets en datascience avec des étudiants en apprentissage, ont démontré l'intérêt d'une animation plus forte sur le thème des données au cœur de Bercy, et par conséquent l'intérêt des données dans nos missions, comme le démontre également l'expérience historique de nos directions.**



**Cette feuille de route actualisée 2020/2022 crée un fil d'Ariane, qui permet à toutes les directions et services de Bercy de s'engager sur un chemin commun.** Je salue les nombreuses actions envers les agents du MEFR, qui sont par nature au cœur de la transformation, en misant sur le développement de leurs compétences. La feuille de route appelle également à plusieurs investissements, comme le développement actif du service public de la donnée, des hub de données sectoriels et les usages innovants de la donnée 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, sans compétence, sans méthode commune dans ce labyrinthe technologique, la réussite est compromise. Cette feuille de route y remédie et liste 25 actions à l'échelle ministérielle. La participation de toutes les directions, des services, des agents, des managers et des cadres dirigeants, est cruciale pour positionner le MEFR dans le siècle des données, des algorithmes et des codes sources.

**Par la valeur qu'elle détient, la donnée a vocation à nourrir les transformations et la modernisation de l'État,** dans les territoires comme en administration 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 ».

Le ministre de l'économie, des finances et de la relance  
Bruno LEMAIRE

## 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, prévisions... mais aussi référentiel des entreprises, code officiel géographique, nomenclature d'activités française... Le ministère dispose d'une immense richesse de données, d'algorithmes et de codes sources tant par sa diversité que par sa profondeur et son expérience.

**La donnée est au cœur des transformations de l'action publique.** Un fonctionnement interne amélioré porteur d'efficacité, à la mise en place de nouveaux services numériques performants, 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 son droit de contrôle des politiques publiques.

La donnée ne doit plus être perçue seulement comme un élément 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.

Les algorithmes et les codes sources participent à la transparence, voire l'acceptation, des décisions de l'administration et promettent un meilleur dynamisme de la recherche, le développement de nouveaux usages et l'innovation, comme l'intelligence artificielle.

Pour fixer un cadre de travail commun au sein de Bercy, autour du triptyque donnée / algorithmes / codes-sources, **la feuille de route ministérielle 2020/2022 dresse 25 actions concrètes** et organisées en trois dimensions liées entre elles.

### 1. 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** sont 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. Architecture

Les enjeux imposent un cadre structuré avant l'exploration du champ des possibles : une organisation claire, des définitions métiers précises et partagées, une connaissance du patrimoine des données, et en particulier les 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 la période pour **gouverner, connaître, partager ,ouvrir les données**, pour partager les algorithmes et échanger les codes sources.

### 3. Usages

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

La mise en place d'un **data lake interne** dédié aux données d'entreprises pour améliorer la coordination de nos services, l'animation d'une **communauté de datascientists** pour partager entre pairs les travaux en cours, réussites ou échecs, et susciter de meilleurs cas d'usages grâce aux données et à 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 trois années.

Les 25 actions sont **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

Une donnée, un algorithme ou un code source ne doivent plus être perçus seulement comme un élément unitaire d'un 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.

Avec la circulaire<sup>1</sup> du Premier ministre signée le 27 avril 2021, le périmètre des AMD est étendu aux algorithmes et codes sources (AMDAC) et la dynamique ministérielle renforcée avec un pilotage stratégique interministériel.

Pour 2020-2022, l'AMDAC 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 :

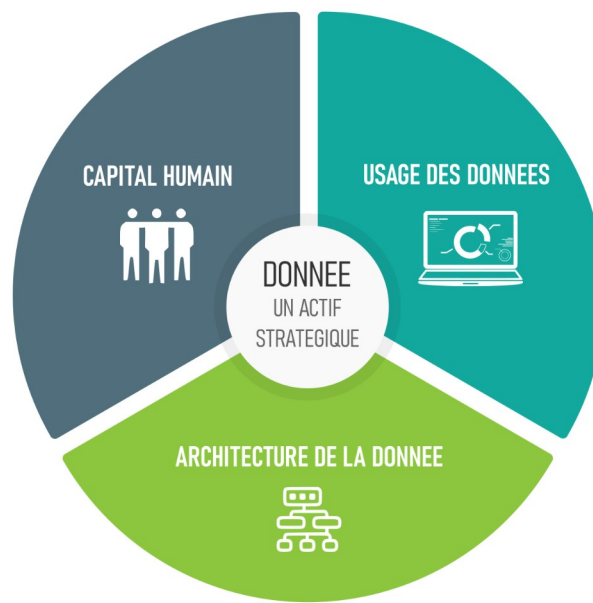
- la dimension « capital humain » : avec une montée en connaissance des agents du MEFR pour qu'ils puissent mobiliser les données dans leurs missions ou comprendre comment les données impactent leur travail de tous les jours ;
- la dimension « architecture de la donnée » : avec une gouvernance, qui incarne la dynamique de transformation au plus haut niveau dans toutes les composantes

1 Circulaire n°6264



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

Le Secrétariat général anime différentes communautés d'experts dans les domaines portées par la feuille de route – datascience, données, données ouvertes, innovations numériques – nécessaires à la sécurisation des travaux, projets et investissements des directions et des services.

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 exigeants 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 « données » 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.....	7
I. Capital humain.....	11
Développer une culture commune qui mêle donnée, Algorithmes et codes sources.....	11
Créer une cohérence entre les métiers, les formations et le parcours professionnel.....	12
II. Architecture.....	14
Gouverner.....	14
Connaître.....	15
Partager et ouvrir les données et les codes sources.....	18
III. Usages.....	20
Proposer des outils de première nécessité.....	20
Favoriser la science de la donnée.....	21
Accompagner les usages innovants.....	22
Annexes.....	23
Planning des actions.....	23
Réseau des administrateurs de données.....	25
abréviations.....	26

## 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 QUI MÊLE DONNÉE, ALGORITHMES ET CODES SOURCES

- **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, aux innovations, aux algorithmes et aux codes sources : des **conférences-débats** mensuelles et, une fois par an, la « journée de la donnée » des ministères économiques et financiers.

Par ailleurs, des **séquences d'information et de sensibilisation** sont également proposées chaque mois pour aider les agents à appréhender les enjeux portés par les données, les algorithmes et les codes sources. 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 ces domaines ou non. Ces séquences doivent permettre de placer la donnée et l'innovation dans la chaîne de valeur d'élaboration d'un service public et d'inspirer de nouveaux usages des données au sein d'un service.

#### Les journées de la donnée

Le SG du MEFR a organisé en 2019 et 2020 une journée dédiée à la donnée, qui a réuni à chaque édition plus de 300 participants. Plusieurs grands témoins se sont succédés, comme Luc Julia directeur numérique de Samsung ou les députés Jean-Noël Barrot et Eric Bothorel. Ces journées permettent des échanges de point de vue lors de tables rondes sur des innovations comme l'intelligence artificielle au service de nos missions ou les compétences à détenir par les experts et aussi par les agents, dont les métiers peuvent évoluer. Enfin ces journées mettent en avant les projets et les initiatives des directions pour illustrer le dynamisme de Bercy et inspirer les autres directions. Ces journées sont ouvertes aux agents publics de Bercy et d'ailleurs.

- **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, nécessaires aux traitements des données. Ces moments de partage d'expertise permettent d'approfondir les connaissances, de valoriser des talents internes et de détecter des innovations pour les expérimenter et ainsi transformer les métiers. Ces temps concernent tout particulièrement des innovateurs des directions.



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

Ces actions expérimentent un nouveau cadre de la gestion des compétences et des ressources humaines pour donner de la clarté 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 avant tout d'identifier les besoins en compétence « données », encore trop rare, au sein des directions de Bercy, et de construire des parcours de formations adaptés.

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

*Datascientist* est aujourd'hui le métier le plus attractif, en particulier par les salaires proposés, 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 est nécessaire 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, en partie déjà coproduit avec la communauté ministérielle, apporte 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 et la DB pour les questions financières.

### 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 renforcer le vivier de compétences internes** (action 4) (CONT)

Selon le Gartner<sup>2</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 très vite et il est nécessaire de s'y préparer.

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. Cela crée 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.

Avec l'IGPDE, il s'agit de faire évoluer chaque année les formations proposées sur le thème des données aux agents spécialistes ou non du domaine, et de les organiser pour créer des parcours pédagogiques adaptés aux métiers et attentes du MEFR. La certification, qui a été testée avec

<sup>2</sup> Source : 2019 Gartner CEO and Senior Business Executive Survey

succès en 2021 auprès de 24 agents publics sur trois formations spécifiques sur le langage open-source R, doit être appréciée comme un moyen de valoriser les nouvelles compétences des agents. L'intégration de la certification dans l'offre IGPDE est à instruire courant 2022.

- **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 plateformes dédiées, comme [DataScientest.com](https://www.data-science-test.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 appréciée des experts publics 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 à la transparence des algorithmes, aux différents secrets et au RGPD seront également à proposer pour développer une meilleure connaissance des principales règles qui régissent les traitements de données, et notamment la protection des données personnelles.

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. L'offre de formations doit être adaptée par défaut au télétravail et peut prendre plusieurs formes, comme des master-class.

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.

**UtilitR : une documentation en ligne pour toute personne intéressée par le traitement des données en R**

Cette documentation en français, rédigée par des utilisateurs de R, propose un panorama des possibilités offertes par ce logiciel très utilisé dans la statistique publique. De nombreuses fiches pratiques aident les utilisateurs à réaliser leurs traitements.

<https://www.utilitr.org/>



## II. ARCHITECTURE

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

### GOUVERNER

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

- **Intégrer la donnée, les algorithmes et les codes sources dans la gouvernance stratégique** de Bercy (action 6) (20S1)

La maîtrise de la donnée sur tout son cycle de vie est un des enjeux du ministère de l'économie, des finances et de la relance tant les usages sont nombreux et créateurs de valeur. À l'instar de la Direction interministérielle du numérique, qui encourage<sup>4</sup> la mue des directions des systèmes d'information ministérielles en direction du numérique<sup>5</sup> et qui installe<sup>6</sup> un comité interministériel des administrateurs des données (CIAD), le pilotage de la donnée, des algorithmes et des codes sources se décline dans la comitologie stratégique et opérationnelle de Bercy.

Par son rôle de coordinateur, le SG a intégré la donnée dans la gouvernance ministérielle, dont la réunion mensuelle des directeurs des systèmes d'information et le comité stratégique des systèmes d'information sous la présidence de la Secrétaire générale et en présence des directeurs d'administration centrale. Plusieurs chantiers liés aux données y sont validés chaque semestre. Les codes sources et les algorithmes sont des sujets qui seront systématisés à l'instar du domaine données.

Par ailleurs, en lien avec le chantier « éclairer les politiques publiques par la données » du programme Tech.Gouv de la DINUM, 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.

- **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.

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

5 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>

6 Circulaire du Premier ministre

Animée par le SG, cette communauté se réunit trois à quatre fois par an et assure le pilotage de la feuille de route « donnée, algorithmes et codes sources » de Bercy. D'autres réseaux – opendata, innovations numériques – sont également animés à un rythme trimestriel (ou autant que de besoin).

- **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. Les différents engagements pris lors des CITP, en particulier celui concernant l'accès au revenu fiscal de référence par des administrations autorisées, illustrent la volonté du ministère à simplifier les démarches par le partage de données.

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 assure une coordination de ce chantier dans un esprit de simplification de l'exposition des API en prenant en compte les spécificités des données. Il a produit des premiers éléments de doctrine en la matière à échanger dans un cadre interministériel. 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 PISTE, la solution de management d'API opérée par l'AIFE, car avec près de 15 millions d'appels par jour, elle propose un niveau de service industriel et elle est retenue par plusieurs de nos partenaires pour publier leurs API. Au titre du FTM, 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 fournisseurs de services à les utiliser.

- **Poser un cadre d'usage du cloud pour les données (action 9) (21S2)**

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. 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 et la plateforme « Onyxia » de l'Insee.

## CONNAÎTRE

Bien avant l'utilisation des données, des codes sources ou des algorithmes, il est nécessaire de développer une vision la plus exhaustive possible et partagée du patrimoine détenu au sein de Bercy. Dans le rapport Foulleron, 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 ainsi lutter contre cette prolifération incontrôlée des données, en premier lieu en partageant largement les référentiels. Les codes sources et algorithmes répondent à la même problématique.



Une cartographie des données, des codes sources et des algorithmes, librement accessible et à jour, qui permet le classement, l'organisation et l'écoulement de ces trois objets selon des critères partagés et avec la prise en compte de la dimension métier. C'est également un chantier relativement chronophage qui nécessitera une mise en œuvre pragmatique.

- **Développer le service public de la donnée (action 10) (21S2)**

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 est également nécessaire pour identifier de nouvelles données d'intérêt à intégrer dans le SPD, avec l'expertise et le soutien de la Dinum, 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		
À ce jour, neuf jeux de données, qui couvrent un large champ thématique ont été identifiés comme des données de référence.		
<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	<b>Base Sirene des entreprises et de leurs établissements (SIREN, SIRET)</b> Pour vous abonner à notre lettre d'information Sirene open data actualisée, suivez ce lien : <a href="https://insee.fr/fr/information/1405555">https://insee.fr/fr/information/1405555</a> Pour	<b>Code Officiel Géographique (COG)</b> Chaque année, l'Insee met à disposition sur son site ( <a href="https://insee.fr">insee.fr</a> ) le code officiel géographique qui rassemble les codes et libellés des communes.
<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	<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	<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
<b>Référentiel à grande échelle (RGE)</b> L'Etat a confié à l'IGN le développement du référentiel à grande échelle (RGE). Pour ce faire, il fait appel à ses moyens propres ainsi	<b>Répertoire National des Associations (RNA)</b> Que contient le RNA ? Le RNA repertorie l'ensemble des associations relevant de la loi du 1er juillet 1901 relative au contrat d'association.	<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

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) (21S2)**

La normalisation et les standards précisent des règles d'écriture des données afin de faciliter les échanges entre les systèmes d'information<sup>7</sup>. Cette démarche favorise l'interopérabilité, l'échange de données et la création de nouveaux services pour les agents du MEFR et les usagers.

<sup>7</sup> Cf. guide ISO/CEI 2 : <https://www.iso.org/fr/standard/39976.html>



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. L'expérience sur les données essentielles de la commande publique sera utile, comme le service [schema.data.gouv.fr](https://schema.data.gouv.fr).

- **Référencer de manière pragmatique les données détenues**, en priorité les référentiels, par les directions et les services (action 12) (21S1)

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. Les travaux pourront également s'inspirer des avancées d'autres ministères et des orientations de la Dinum.

- **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 pour les agents ou les entreprises et facilite les échanges de données entre les différentes applications.

D'autres directions de Bercy, comme la DGFIP ou la DGDDI, utilisent également 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.finances.gouv.fr](https://terminologie.finances.gouv.fr) ; il sera également et prochainement enrichi de nouveaux termes numériques et données, comme *datalake*, *dataops*, *dataviz*, API, pile IA, *tracking*, *tracing*, etc.



Il est attendu que chaque direction se positionne sur un jeu de données en cohérence avec son programme de travail de 2021 et 2022.

### Identité numérique des agents, entreprises et particuliers

Pour améliorer la sécurité des accès aux données et aux systèmes d'information, le SHFDS propose la carte d'identité numérique « rossignol ». Car l'identité professionnelle et la maîtrise des modèles des droits d'accès aux applicatifs sont des sujets données les plus communs dans un système d'information. Des projets de gestion des identités, comme France-Connect, répondent à ce défi de maîtrise des référentiels, thème central au sein du MEFR.

### 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 des marchés publics et des contrats de concession qu'ils attribuent. À moyen terme, les données de recensement seront 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 par ex.). En associant les éditeurs à une stratégie de la donnée, ce projet assure une cohérence des données en lien avec les usages qu'elle permet par son ouverture.

## PARTAGER ET OUVRIR LES DONNÉES ET LES CODES SOURCES

Dans la continuité des conclusions du rapport Bothorel et des engagements du CITP, le MEFR poursuit son action vers plus de valorisation des données publiques, des codes sources ou des données, déjà partagés ou nouvellement ouverts, facilitant l'innovation ou évitant aux citoyens de fournir plusieurs fois les mêmes documents aux administrations. La plateforme ministériel [data.economie.gouv.fr](http://data.economie.gouv.fr) facilite l'accès aux données ouvertes de Bercy<sup>8</sup> et s'intègre parfaitement à l'écosystème national [data.gouv.fr](http://data.gouv.fr). L'accès à certaines données, comme le revenu fiscal de référence, permet à de nombreux partenaires – administrations ou collectivités territoriales – d'engager une simplification de leurs démarches administratives<sup>9</sup>.

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

Le processus du cycle de vie de la donnée<sup>10</sup> couvre un champ assez large : la collecte, le stockage, le traitement, la valorisation et l'archivage des données. Un processus a été défini en collaboration avec la communauté ministérielle, 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. Ce processus est partagé avec la Dinum et les autres ministères.

- **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

### 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 l'unicité d'inscription grâce à un travail inédit de mise en qualité des données.

8 À l'exception des données statistiques de l'Insee largement diffusées sur son site institutionnel parfaitement identifié

9 Par exemple : inscription à la cantine, demande d'aide juridictionnelle, demande de logement social, aide à la rénovation énergétique des logements

10 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.

inexploitables, 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 l'expérience acquise sur des données reconnue, comme le répertoire Sirene de l'Insee ou les demandes de valeur foncière (DVF) de la DGFIP, 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.

Les directions et services du MEFR s'engagent à documenter un à deux jeux de données par an.

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

L'ouverture des données impose, sauf avis contraire du législateur<sup>11</sup>, la protection des données personnelles et par construction l'impossibilité de pouvoir y retrouver un individu 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>12</sup>) et la Dinum, un groupe de travail sera mis en place pour partager les bonnes pratiques et apporter conseils et exemples au producteur de données.

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

L'efficacité des services et des directions peut se mesurer en partie grâce à leur capacité d'échanger et partager leurs données, leurs codes sources et leurs algorithmes au sein de la communauté ministérielle ou interministérielle. Des demandes, en particulier celles de complexité importante, peuvent nécessiter un suivi spécifique. Une chambre d'enregistrement et de suivi des demandes est mise en place. Elle assure une mutualisation des questions juridiques au regard du droit de la donnée ou de la propriété parfois complexe. Des premières demandes ont d'ores et déjà été exprimées par des directions générales, comme celle de la CCRF pour des besoins de profilage sur la question de la fraude. Cette chambre traite également des demandes externes.

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

Avec plus de 365 données ouvertes par les directions et services du MEFR sur le portail [data.economie.gouv.fr](https://data.economie.gouv.fr), de nombreuses statistiques publiques publiées par l'Insee et des dizaines de dépôts pour les codes sources sur [GitHub](https://github.com) ou d'autres plateformes<sup>13</sup>, le ministère démontre son engagement et son dynamisme dans la politique d'ouverture et de transparence. Bien que de nouvelles données seront publiées, comme le ministère s'y engage lors des comités interministériels de la transformation publique<sup>14</sup>, les efforts doivent maintenant se concentrer sur les usages de ces données mises à disposition de tous, car elles sont encore insuffisamment connues et mobilisées par la « communauté opendata ». Pour y remédier, le SG du MEFR initie une communauté qui vise à rapprocher les producteurs et utilisateurs de données et de codes sources de Bercy.

Un rendez-vous, de type « non conférence », est proposé en novembre de chaque année et donne la parole à la communauté « Bercy Ouvert ». Dans la continuité des deux hackathons DataFin organisés par la DGFIP, la DB et la Dinum, d'autres événements seront également proposés en lien avec les directions et services de Bercy.

11 Par exemple : données de décès, données des permis de construire

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

13 Par exemple : code source d'OPALE, de la calculatrice des impôts, codes sources de startup d'État, code source pour le traitement des données essentielles de la commande publique (DECP-rama)

14 Parmi les engagements : codes sources du calcul de la cotisation foncière des entreprises (CFE), du calcul de la cotisation à la valeur ajoutée des entreprises (CVAE), du prélèvement à la source et ouverture des demandes de valeurs foncières pour l'Alsace-Moselle et des indicateurs d'usage de Signal Conso et de Rappel Conso



### III. USAGES

*Outiller, valoriser, mêler*

Les données ne s'usent que lorsque l'on ne s'en sert pas... Leur mobilisation porte de nombreuses simplifications au bénéfice des agents ou du public, dont la simplification des démarches administratives ou à l'administration proactive, deux chantiers particulièrement symboliques du partage automatique de données entre acteurs.

#### PROPOSER DES OUTILS DE PREMIÈRE NÉCESSITÉ

- **Expérimenter la mise en place d'un datahub 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 datahub s'impose comme un outil incontournable d'une stratégie de la donnée.

Pour les missions de pilotage budgétaire ou de contrôle 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 précédente, avec une expérience utilisateur personnalisée et avec des données collectées en temps réel. Le datahub 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 nos nombreux métiers.

Au 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, comme la DGE et le datahub des aides de l'État, 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 à résoudre, en privilégiant l'opensource par un partenariat avec des acteurs comme le [TOSIT](#), un 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 secrets et la protection des données personnelles.

- **Prototyper une architecture « data-ops »** (action 20) **(22S1)**

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>15</sup>) sont à proposer pour créer des automatismes sur le cycle de vie de la donnée.

15 L'objectif est de proposer aux cadres dirigeants 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

## 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 des 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 les intégrer dans leur programme de recrutement.

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.

Toutes les directions n'ont pas les mêmes moyens pour y parvenir. Il est nécessaire de rassembler les forces<sup>16</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 des outils** mis en œuvre au 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.

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.

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



## ACCOMPAGNER LES USAGES INNOVANTS

- **Proposer une expertise datascience** (action 24) (21S1)

Les projets, qui nécessitent de la datascience, sont à accompagner avec volontarisme et coordination entre les directions grâce à un apport en expertise. Le SG met en place quelques ressources partagées dans le domaine de la science des données et de l'intelligence artificielle pour accompagner les directions dans leurs projets. Ces ressources, principalement recrutés dans un vivier d'apprentis encadrés par le SG, pourront apporter leur expertise aux directions qui en feront la demande sur un projet en cours ou à venir.

- **Favoriser l'innovation sous forme de startup d'État dans l'incubateur 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 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 l'organisation du ministère, le mode start-up est encouragé par le Secrétariat général reste à l'initiative des directions et des services, qui pourront alors mobiliser l'incubateur ministériel qui fournit cadre et support, le tout inspiré de l'incubateur national.

## ANNEXES

### PLANNING 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 renforcer le 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 et des codes sources** (action 18)
- **Expérimenter des innovations autour de l'IA** (action 22)
- **Favoriser l'innovation sous forme de startup d'État dans l'incubateur du MEFR**, membre du réseau Beta.gouv (action 25)

#### Les actions du second semestre 2020

- **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 des outils mis en œuvre au MEFR** (action 23)

#### Les actions du premier semestre 2021

- **Référencer de manière pragmatique les données détenues**, en priorité les référentiels, par les directions et les services (action 12)
- **Mettre en place une expertise datascience** (action 24)





#### Les actions du second semestre 2021

- **Poser un cadre d'usage du *cloud* pour les données** (action 9)
- **Développer le service public de la donnée** (action 10)
- **Définir une norme de la donnée** (cf. schéma) (action 11)
- **Expérimenter la mise en place d'un datalake interne dédié aux données d'entreprises** (action 19)

#### Les actions du premier semestre 2022

- **Urbaniser des données** (action 13)
- **Prototyper une architecture « data-ops »** (action 20)

#### Les actions du second semestre 2022

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



## RÉSEAU DES ADMINISTRATEURS DE DONNÉES

- **AIFE** : Laurent Robillard
- **APE** : Franck Boiteau
- **CBCM** : Philippe Sauvage
- **CGEF** : Christian Avazeri, Luc Degardin
- **CGEJET** :-
- **DAJ** : Serge Doumain
- **DB**: Vincent Amilhaud
- **DGCCRF** : Guillaume Defillon
- **DGDDI** : Christophe de Saint-Viance
- **DGE** : Raphaël Keller
- **DGFIP (yc DIE et DLF)** : Lionel Ploquin
- **DG INSEE** : Thierry Aouizerate, Loïc Midy
- **DGTRESOR** : Antonin Aviat, Yannis Kemel
- **IGF** :-
- **MICAF** : Christine Fournier
- **SG** : Stéphane Trainel (administrateur ministériel des données, des algorithmes et des codes sources)
- **TRACFIN** : Delphine Le



## 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

AMDAC : Administrateur ministériel des données, des algorithmes et des codes sources

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

CIAD : Comité interministériel des administrateurs des données

CITP : Comité interministériel de la transformation publique

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 publique

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

MICAF : Mission interministérielle de coordination anti-fraude

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'État

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'open source





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

*Liberté  
Égalité  
Fraternité*