



**MINISTÈRE  
DE L'ENSEIGNEMENT  
SUPÉRIEUR,  
DE LA RECHERCHE  
ET DE L'INNOVATION**

*Liberté  
Égalité  
Fraternité*

# Politique des données, des algorithmes et des codes sources

Feuille de route  
2021-2024




## Préface

La crise sanitaire que nous connaissons depuis plus d'une année a démontré que la donnée constituait un enjeu majeur en période de gestion de crise, pour éclairer la décision publique et informer le citoyen dans un souci de transparence par rapport aux décisions prises. L'ouverture des données des ministères et des agences régionales de santé a aussi favorisé les initiatives citoyennes et ainsi alimenté le débat public. Cette période a été le révélateur de l'intérêt des citoyens pour la recherche et la publication d'indicateurs en temps réel. Plus que jamais, la donnée s'est trouvée au cœur de la décision et du pilotage des actions et a montré l'importance, à l'échelle planétaire, de l'ouverture des connaissances scientifiques. L'ouverture et une bonne circulation des données est un travail de fond qui mobilise aussi bien l'administration publique, dont la recherche, que le secteur privé. La structuration, la circulation et l'ouverture des données de l'enseignement supérieur, de la recherche et de l'innovation, initiées depuis plusieurs années, notamment avec le mouvement de la science ouverte, doit désormais se généraliser à tous les types de données, et s'accélérer pour devenir un levier majeur de création de savoirs et de valeurs économique, démocratique, scientifique et politique.

Dès 2016, le législateur a perçu l'intérêt des données en promulguant la loi pour une République numérique demandant à l'administration publique d'être plus engagée dans l'ouverture de ses données et codes sources. La loi n° 2018-166 du 8 mars 2018 relative à l'orientation et à la réussite des étudiants s'est inscrite dans cette dynamique tout en recherchant un équilibre entre les intérêts publics en cause. Il en est résulté notamment la publication du cœur algorithmique de l'application *Parcoursup* et la publication régulière sur *dataESR* d'un ensemble de données au titre de l'évaluation de la loi et de la transparence publique. Le *Plan national pour la science ouverte* lancé en 2018 par la ministre de l'Enseignement supérieur, de la Recherche et de l'Innovation a permis à la France de se doter d'une politique cohérente et dynamique en faveur de l'ouverture ou le partage des données et codes sources de la recherche.

La nouvelle politique de la donnée voulue par le Premier ministre<sup>1</sup> constitue une priorité stratégique de l'Etat et conduit le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation à renforcer les efforts déjà engagés en déployant une nouvelle feuille de route engagée et ambitieuse qui, d'une part, fixe les enjeux de la politique des données, des algorithmes et des codes sources et, d'autre part, met en place un plan d'actions à 4 ans concernant les données et les codes sources de l'administration, de l'enseignement supérieur, de la recherche et de l'innovation. Cette feuille de route s'articule avec les engagements pris dans le cadre d'un Partenariat pour un gouvernement ouvert et avec le deuxième Plan national pour la science ouverte. Grâce aux initiatives et à la mobilisation de ses services et de ses opérateurs, la diffusion et l'ouverture des données, des algorithmes et des codes sources constitueront, pour l'enseignement supérieur, la recherche et l'innovation, un véritable levier pour renforcer l'efficacité et la transparence de leurs actions et accélérer les découvertes scientifiques.



**Claire GIRY**

Directrice générale de la recherche et de l'innovation



**Anne-Sophie BARTHEZ**

Directrice générale de l'enseignement supérieur et de l'insertion professionnelle

<sup>1</sup> Voir Circulaire du premier Ministre n° 6264-SG du 27 avril 2021

## Résumé

### *Circulation, ouverture et réutilisation des données au cœur de la transformation de l'action publique et de l'accélération des découvertes scientifiques*

Le potentiel des données d'enseignement supérieur et de recherche, ainsi que des données administratives et de pilotage de l'enseignement supérieur, de la recherche et de l'innovation est encore trop peu exploité, malgré les efforts déjà engagés. Pourtant, une meilleure gestion des données permettrait un gain important de valeur scientifique, économique et d'efficacité collective.

Dans le cadre des textes législatifs et réglementaires en vigueur et dans le prolongement des orientations interministérielles susmentionnées, la feuille de route du MESRI fixe un cadre de travail commun au ministère et à ses opérateurs pour répondre à trois enjeux majeurs :

- Développer l'ouverture des données et des codes sources logiciels pour permettre leur réutilisation par tous et favoriser l'**innovation**, également créatrice de valeur scientifique et économique ;
- Amplifier la transparence, *via* l'ouverture et l'accès aux données, aux algorithmes et aux codes sources pour davantage de **confiance dans l'action publique** ;
- Améliorer la circulation et l'utilisation des données pour la **simplification** des processus et l'aide au pilotage.

Pour porter ces enjeux, le MESRI a adopté une démarche s'appuyant sur l'ensemble des expertises et des initiatives déjà existantes au sein du ministère et chez ses opérateurs. La feuille de route 2021-2024 rassemble, capitalise, promeut, amplifie et mutualise les initiatives répondant aux attentes des usagers. Cette feuille de route témoigne d'une volonté politique mais surtout, elle détaille les objectifs qui se déploient en 53 actions concrètes.

**Compétences** • La valeur de la donnée réside dans son usage. L'usage de la donnée et les innovations numériques impactent les métiers et l'évolution des compétences qui sont centraux dans la réussite de la démarche. Développer et reconnaître les compétences en matière de gestion, traitement, préservation et ouverture des données des étudiants, des enseignants-chercheurs, des chercheurs, comme des personnels d'appui à la recherche ou de l'administration tout au long de leur parcours est une priorité.

**Gestion et réutilisation des données** • Les actions visent tout d'abord à améliorer la gestion des données de la recherche, de l'enseignement supérieur et de pilotage, afin d'amplifier leur ouverture ou leur partage et leur réutilisation. Il s'agit de développer et mutualiser les services visant à garantir l'accès et l'exploitation de ces données par tous et à renforcer notre souveraineté.

**Codes sources logiciels et usage des logiciels libres** • Une politique et des services de préservation et d'ouverture des algorithmes et des codes sources seront mis en œuvre. En parallèle, le principe de priorité d'usage des logiciels libres sera réaffirmé afin de garantir des solutions accessibles à tous et favoriser la souveraineté numérique française et européenne.

**Simplification** • Des mesures de simplification et d'harmonisation des processus de candidatures des chercheurs à des appels d'offre et des étudiants aux formations de master seront mises en œuvre.

**Transparence** • L'ouverture des données d'indicateurs du ministère et de ses opérateurs sera favorisée afin de participer à l'évaluation et au suivi de l'impact des politiques publiques mises en place par le MESRI et ses opérateurs.

La feuille de route sera suivie par les deux directrices générales du MESRI, afin d'évaluer son avancement, adapter sa trajectoire aux besoins, priorités ou opportunités et mesurer les impacts tangibles dans la transformation publique et l'évolution des pratiques d'enseignement et de recherche.

# Sommaire

I.	Contexte et opportunités .....	7
II.	Enjeux et objectifs .....	9
III.	Plan d'actions en faveur de l'ouverture, du partage et de l'exploitation des données, des algorithmes et des codes sources .....	11
A.	Ouverture, partage et exploitation des données.....	11
	Rendre les données de l'enseignement supérieur, de la recherche et de l'innovation disponibles et accessibles - <i>#dataESR</i> .....	12
	Augmenter la visibilité de la recherche et de l'innovation française – <i>ScanR</i> .....	12
	Valoriser et accroître les ressources libres de l'enseignement supérieur – <i>Canal U</i> .....	13
	Enrichir le service de formations en ligne pour les étudiants - <i>FUN</i> .....	13
	Aider à la création de formations en ligne - <i>Callisto</i> .....	14
	Structurer, ouvrir et partager les données de recherche - <i>RechercheDataGouv</i> .....	14
	Suivre l'ouverture des données et des codes de la recherche - <i>Baromètre de la science ouverte</i> .....	15
B.	Gestion des données de la recherche .....	15
	Définir une stratégie nationale de services numériques dédiés à la préservation des données de la recherche .....	16
	Accompagner les chercheurs dans la gestion des données et le .....	16
	« FAIR <i>by design</i> » .....	16
C.	Préservation et ouverture des algorithmes et des codes sources.....	16
	Déposer les codes sources publics du MESRI sur <i>code.etalab.gouv.fr</i> .....	16
	Collecter, préserver et partager les codes sources – <i>Software Heritage</i> .....	17
	Poursuivre l'ouverture des algorithmes et des données - <i>Parcoursup</i> .....	17
D.	Usage des logiciels libres .....	17
	Privilégier l'usage des logiciels libres .....	18
	Encourager la production et l'usage de logiciels libres – <i>esup portail</i> .....	19
E.	Ecosystème des infrastructures de la science ouverte.....	19
	Développer et animer l'écosystème des infrastructures de la science ouverte .....	19
F.	Simplifier la vie des chercheurs, des enseignants-chercheurs et des étudiants .	20
	Accélérer les demandes d'accès des chercheurs aux données publiques .....	20
	Favoriser l'accès des chercheurs aux données d'intérêt général .....	20
	Mettre en place des identifiants uniques et pérennes .....	21
	Faciliter les candidatures aux appels à projet de recherche .....	22
	Simplifier les démarches pour les étudiants grâce aux API.....	22
	Améliorer la plateforme <i>Trouver mon master</i> .....	22

G.	Impacts des politiques publiques.....	23
	Engager les établissements de recherche et d'enseignement supérieur à créer et diffuser les jeux de données sur leurs résultats et leurs moyens.....	23
	Engager les agences de financement à créer et diffuser les jeux de données présentant les projets lauréats et leurs financements.....	23
	Intensifier la transparence des dépenses d'abonnements.....	23
	Améliorer l'usages des services numériques de formation.....	23
	Créer des données sur les usages des infrastructures de la science ouverte.....	24
	Elargir le lectorat des articles scientifiques grâce à la traduction.....	24
IV.	Maîtrise des données et des codes sources : acculturation, accompagnement, développement et reconnaissance des compétences.....	25
A.	Actions au sein du MESRI.....	25
	Mettre en place des « rendez-vous » de la donnée.....	25
	Former aux métiers de la donnée.....	26
	Produire des données de meilleure qualité.....	26
	Éditorialiser et rendre intelligible les données pour favoriser leur réutilisation par certains publics.....	26
B.	Accompagner la structuration, le partage et l'ouverture des données et codes sources de la recherche.....	26
	Structurer l'offre des « Ateliers de la donnée ».....	27
	Soutenir et participer à la <i>Research Data Alliance (RDA)</i> .....	28
	Accompagner l'adoption de licences libres.....	28
C.	Créer une cohérence entre métiers, formations et parcours.....	28
	Introduire une politique de formation initiale à l'usage des données et des codes sources.....	29
	Reconnaître la diversité des productions et des contributions à la recherche.....	29
	Reconnaître et accompagner l'évolution des carrières.....	29
V.	Piloter la feuille de route et évaluer son impact.....	30
A.	Principes directeurs de la feuille de route.....	30
B.	Administrateur ministériel des données, des algorithmes et des codes sources.....	30
C.	Assurer la mobilisation au sein du ministère et développer les synergies.....	31
	Cataloguer et décrire les données pour dynamiser les synergies de partage.....	31
	Créer une charte des rôles et responsabilités du capital des données du MESRI... ..	32
D.	Un réseau d'administrateurs des données, des algorithmes et des codes sources de la recherche.....	32
	Constituer et animer le réseau des administrateurs des données, algorithmes et codes sources de la recherche.....	32
E.	Une gouvernance adaptée au pilotage de la feuille de route.....	32
F.	Mesure des impacts des actions de la feuille de route.....	33
	Actions faisant l'objet de dépôts de candidature au plan de relance.....	34

# I. Contexte et opportunités

***Spécificité des données, algorithmes et codes sources du MESRI*** : données, algorithmes et codes sources qu'il s'agisse des données administratives, de pilotage, d'enseignement, de recherche ou d'innovation. Les données et les codes sources sont détenus par le Ministère et la diversité de ses opérateurs d'enseignement supérieur et de recherche. Les données et codes sources de la recherche et de la formation sont détenus par les équipes de recherche et d'enseignement supérieur. Plus de 800 000 personnels répartis dans plus d'une centaine d'établissements contribuent à l'acquisition, la gestion, la préservation et l'ouverture des données et des codes sources.

L'élaboration de la feuille de route du MESRI s'inscrit dans l'élan de transformation français, « pour plus d'ouverture de données », mais aussi européen : « La stratégie européenne en matière de données doit permettre [ ... ] la création d'un marché unique des données qui favorisera la libre circulation de ces dernières au sein de l'UE et entre les secteurs, dans l'intérêt des entreprises, des chercheurs et des administrations publiques ».<sup>2</sup> De dimension française et européenne, cette feuille de route s'approprie les propositions d'actions recommandées par le député Bothorel<sup>3</sup> sur l'ouverture des données et du code logiciel. Elle s'inscrit aussi dans les efforts de la France dans le cadre de la transparence de l'action publique, via l'OGP – *Open government partnership* et prend sa part d'engagements concernant les mesures en faveur de l'ouverture et du partage des données et codes sources de la recherche inscrites dans la politique nationale de la science ouverte<sup>4</sup>. Ces différents engagements et politiques portent une stratégie globale et cohérente en matière d'ouverture et de transparence de l'action, en ce sens, ils se répondent et se complètent.

En dépit des ambitions en matière de science ouverte affichées par le G7<sup>5</sup>, par l'OCDE<sup>6</sup>, par l'UNESCO<sup>7</sup> ou par la Commission européenne qui vient d'en faire un critère d'excellence scientifique et bien sûr par les cinq pays européens, dont la France, qui se sont dotés d'un *Plan national pour la science ouverte*, **le potentiel scientifique et économique des données de recherche est encore insuffisamment mobilisé.**

Bien que le MESRI se soit engagé dès 2013 dans l'ouverture des données relatives à son action et à celle de ses opérateurs, puis en 2018 au travers du *Plan national pour la science ouverte* dans l'ouverture ou le partage des données et codes sources de la recherche, la tâche reste immense.

Le MESRI souhaite, au travers de cette feuille de route 2021-2024, **placer la donnée et le code source**, d'une part **au cœur de la transformation de l'action du ministère** en faisant de la donnée un actif stratégique et un bien commun à tous les services et d'autre part, **au cœur de la transformation durable des pratiques d'enseignement et de recherche.**

---

<sup>2</sup> Stratégie européenne des données, 19 février 2020. [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy\\_fr](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_fr)

<sup>3</sup> Pour une politique publique de la donnée, décembre 2020, Mission confiée par le premier ministre à Éric Bothorel, député des Côtes-d'Armor, Stéphanie Combes, directrice du Health Data Hub, Renaud Vedel, coordonnateur national pour l'IA.

<sup>4</sup> Deuxième plan national pour la science ouverte, 7 juillet 2021. <https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>

<sup>5</sup> "The G7 Open Science Working Group Action Points – Speeding up Open Science?" <https://www.zbw-mediatalk.eu/2016/11/the-g7-open-science-working-group-action-points-speeding-up-open-science/> et <https://www.ouvrirlascience.fr/evenement-g7-mesri/>

<sup>6</sup> *Recommendation of the Council concerning Access to Research Data from Public Funding* - <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347>

<sup>7</sup> UNESCO Recommendation on Open Science <https://en.unesco.org/science-sustainable-future/open-science/recommendation>

Outre l'inscription de certaines de ses actions dans un cadre européen, le MESRI collabore directement avec d'autres ministères ou leurs opérateurs. Les opportunités ou interactions déjà engagées concernent notamment le Ministère de l'éducation nationale de la jeunesse et des sports au travers de la Direction du numérique éducatif, le Ministère de la Transformation et de la Fonction Publique - MTFP sur tous les sujets relatifs à l'open data et au Partenariat pour un gouvernement ouvert, ainsi que la DINUM dans le cadre du réseau des administrateurs ministériels des données et de la mission pour le logiciel libre.



## II. Enjeux et objectifs

### Une stratégie au service de 3 enjeux

#### Innovation

Ouvrir les données, les algorithmes et les codes sources afin de favoriser leur réutilisation par les chercheurs, les enseignants, les citoyens, les organisations publiques et privées, et la société dans son ensemble

#### Confiance

Amplifier l'ouverture des données, des algorithmes et des codes sources pour plus de transparence de l'action publique

#### Simplification

Faire de la circulation des données un outil pour simplifier et rendre plus efficaces les actions et les processus administratifs

**La démarche pour porter ces enjeux s'appuie sur l'ensemble des initiatives existantes au sein du MESRI mais aussi chez les opérateurs du Ministère** que ce soit les établissements d'enseignement supérieur, les opérateurs de recherche, de financement, d'évaluation...

L'élaboration de cette feuille de route s'est appuyée sur une analyse de l'existant, les retours d'expérience des services de l'administration centrale et des opérateurs, les forces et faiblesses des pratiques et dispositifs en place, les opportunités françaises et européennes, les attentes et enjeux des acteurs dans un souci de meilleure circulation des données, d'efficacité et de mutualisation des efforts.

### Une démarche guidée par 5 objectifs

Promouvoir, développer et mutualiser des services d'ouverture et d'usage des données

Participer à la souveraineté et à l'indépendance technologique française et européenne

Développer la culture et l'usage de la donnée

Faire des données, des algorithmes et des codes un bien commun au service de tous

Améliorer la visibilité internationale des données de l'enseignement supérieur, de la recherche et de l'innovation

### **Promouvoir, développer et mutualiser des services d'ouverture et d'usage des données**

Simplifier l'ouverture ou le partage de données directement par leurs producteurs nécessite la promotion ou le développement de services identifiables et simples d'usage. Les évolutions ou

développements de services techniques, de formation et d'accompagnement se feront dans une **logique de maîtrise des efforts** et s'appuiera donc en priorité sur des **succès existants** en favorisant leur **mutualisation** ou leur passage à l'échelle afin qu'ils bénéficient à tous. La finalité des services en appui à l'ouverture des données et des codes sources est de permettre une **réutilisation active**. Cette dernière sera donc explicitée, guidée et promue.

## Participer à la souveraineté et à l'indépendance technologique française et européenne

Les données et les codes sources détiennent un très fort potentiel de création de valeurs à condition de savoir comment les mobiliser, les partager, les tracer et les réutiliser dans un **cadre de confiance défini et équitable** pour les producteurs et les réutilisateurs. Ouvrir codes sources et données en leur attribuant des licences libres favorise le potentiel de réutilisation par tous, reconnaît la contribution des acteurs et services producteurs et favorise notre **souveraineté et notre autonomie stratégique**. Le développement maîtrisé d'infrastructures support pour les données et les codes sources constitue également un levier majeur de cette souveraineté. Le MESRI porte cette stratégie au travers de la feuille de route nationale rassemblant une centaine d'infrastructures de recherche<sup>8</sup>. Ces infrastructures sont nationales ou font l'objet de partenariats internationaux ou européens, notamment par leur engagement dans la feuille de route du forum stratégique européen (ESFRI).

## Développer la culture et l'usage de la donnée

Le MESRI se saisit des recommandations du rapport du député Bothorel engageant l'Etat à définir une **politique d'acculturation et de formation ambitieuse** sur les enjeux de la donnée et plus largement du numérique. Il s'agit d'inscrire ces enjeux dans les cursus de formations de l'enseignement supérieur ainsi que dans la formation continue des personnels de l'enseignement, de la recherche et de l'administration.

## Faire des données un bien commun au service de tous

Les données doivent devenir un bien commun dont chacun est dépositaire et responsable tout en ayant conscience qu'il s'agit d'un actif immatériel stratégique **porteur de valeurs pour tous**. Les données doivent circuler, être partageables entre les différents acteurs de l'ESRI au bénéfice d'une plus grande efficacité collective mais aussi des citoyens et des entreprises.

## Améliorer la visibilité internationale des données de l'enseignement supérieur, de la recherche et de l'innovation

« *La France s'est placée à l'avant-garde européenne de la politique de la donnée et des codes sources depuis 2013, mais l'avance acquise est fragile.* »<sup>9</sup> L'ensemble de l'ESRI déjà engagé dans l'ouverture de ses données et codes sources souhaite activement contribuer au **rayonnement européen et international de la France** en favorisant l'innovation par l'ouverture des données et des codes sources.


---

<sup>8</sup>[https://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures\\_de\\_recherche/70/3/Brochure\\_Infrastructures\\_2018\\_948703.pdf](https://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures_de_recherche/70/3/Brochure_Infrastructures_2018_948703.pdf)

<sup>9</sup> Pour une politique publique de la donnée, décembre 2020, Mission confiée par le premier ministre à Éric Bothorel, député des Côtes-d'Armor, Stéphanie Combes, directrice du Health Data Hub, Renaud Vedel, coordonnateur national pour l'IA.

### III. Plan d'actions en faveur de l'ouverture, du partage et de l'exploitation des données, des algorithmes et des codes sources

Le portefeuille d'actions proposé s'inscrit sur la période 2021-2024 en prenant en compte les enjeux, les attentes des demandeurs et bénéficiaires, ainsi que les ressources pour les conduire. Tout au long du document, les actions sont signalées :

- par le pictogramme 
- référencées [CONT] pour des actions continues
- numérotées au format [A01-AA $S_n$ ] à échéance de l'action : « A » pour action, « 01 » : numéro de l'action, « AA » pour année (ex :22), «  $S_n$  » pour Semestre (ex :S1) - 22S1

#### A. Ouverture, partage et exploitation des données

**L'obligation d'ouverture des données publiques, posée par la loi pour une République numérique de 2016 et organisée par les textes législatifs et réglementaires doit se traduire davantage dans les pratiques**, qu'elles soient celles de l'administration, de l'enseignement supérieur ou de la recherche. Pour cela, l'adaptation ou le développement de plateformes de données et des services d'accompagnement seront mis en œuvre.

L'ouverture de certaines données est limitée par les exceptions légitimes encadrées par la loi, par exemple en ce qui concerne le secret professionnel, la sécurité des systèmes d'information, les secrets industriels et commerciaux, les données personnelles, la protection du potentiel scientifique et technique, les contenus protégés par le droit d'auteur ou par la loi pour l'orientation et la réussite étudiante. Dans ces cas, les pratiques de partage des données devront être favorisées à travers la définition de protocoles maîtrisés.

Notre ambition est de faire en sorte que les données produites dans le cadre de l'administration, de l'enseignement ou de la recherche publique française soient progressivement réutilisables par d'autres que leurs producteurs. Pour cela, les contributeurs de la recherche sont incités et accompagnés dans la structuration des données de recherche en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) afin de les préserver et les ouvrir ou les partager.

Ces principes de description des données par des méta-données, d'accessibilité, d'interopérabilité, de découvrabilité et de réutilisation seront adaptés, lorsque cela est pertinent, pour la structuration des données à caractère administratif. Cela favorisera une meilleure circulation des données et leur réutilisation au sein de l'administration et pour certaines d'entre elles leur ouverture aux citoyens dans le cadre de l'engagement du MESRI en termes de transparence de l'action publique.

La mise en place d'un cadre de bonnes pratiques adaptées à chaque domaine de données a pour objectif de faire mieux circuler les données entre les différents acteurs de l'ESRI afin que les données constituent un bien commun construits par tous et pour tous. Cette démarche se fera au bénéfice d'une plus grande efficacité collective mais aussi des citoyens et des entreprises.

Pour engager chacun des domaines, administration, enseignement supérieur et recherche dans des pratiques actives de description, partage et ouverture de données, les plateformes de diffusion déjà existantes ou en cours de construction seront promues et évolueront en fonction des besoins et des usages. Ces **infrastructures support à l'ouverture et au partage des données** sont développées par et pour la communauté de producteurs de données, **à partir de logiciels dont le code source est publié sous licence libre**. Elles contribuent au **renforcement de la souveraineté**. Cette politique est

également porteuse d'une stratégie de **mutualisation** et de réduction du nombre d'infrastructures et s'inscrit dans une démarche **écoresponsable** du MESRI et de ses opérateurs. Le chapitre IV de cette feuille de route détaille les actions de formation et d'accompagnement pour faire évoluer les pratiques de gestion des données.

## ➤ **Rendre les données de l'enseignement supérieur, de la recherche et de l'innovation disponibles et accessibles - #dataESR**

Confiance, Simplification & Innovation - Action 1 [A01-CONT]

Engagé depuis 2013 dans l'open data, le MESRI contribue à la **transparence des politiques et actions publiques de l'enseignement supérieur, de la recherche et de l'innovation grâce à #data ESR**. La mise à jour des jeux de données et API sera poursuivie. DataESR sera également enrichi par des données, notamment sur la santé et la réussite étudiantes, ou

# #dataESR

#dataESR référence **126 jeux de données open data et API** sur les tableaux de bord de l'enseignement supérieur, la recherche et l'innovation.

[data.esr.gouv.fr](https://data.esr.gouv.fr)

l'engagement associatif. Une **éditorialisation des contenus sera engagée afin de favoriser la découvrabilité des données et ainsi promouvoir leur réutilisation** par les acteurs de l'ESRI comme par les citoyens ou les entreprises. #data ESR met en visibilité les données des établissements en les rassemblant. Les opérateurs du ministère seront encouragés à ouvrir plus leurs données et tout particulièrement sur #data ESR afin d'éviter l'éparpillement des efforts et des plateformes *open data* dans chacun des établissements. #data ESR est développé et maintenu par le département des outils à la décision de la sous-direction des systèmes d'information et des études statistiques.

## ➤ **Augmenter la visibilité de la recherche et de l'innovation française – ScanR**

Confiance, Simplification & Innovation - Action 2 [A02-CONT]



scanR permet d'explorer le paysage de la recherche et l'innovation des laboratoires publics et des entreprises en France parmi un vaste ensemble d'informations reliées entre elles.

### scanR EN CHIFFRES (2021)

42 000 entités publiques et privées  
500 000 fiches auteurs  
110 000 financements  
2 800 000 de publications, thèses et brevets

<https://scanr.enseignementsup-recherche.gouv.fr/>

Créé en 2016, doté d'une nouvelle version en 2020 et enrichi continuellement, **scanR favorise la transparence et l'appropriation par tous des dernières avancées scientifiques**. Il réutilise des données publiques ouvertes, nettoyées et enrichies puis agrégées. Ses **données, librement réutilisables**, sont accessibles à tous *via* les API qui alimentent l'application et grâce aux jeux de données ouverts disponibles sur la plateforme Opendata du MESRI. Son **code source est également sous licence libre**<sup>10</sup>.

Point de visibilité majeur de la recherche et de l'innovation en France et à l'international, son enrichissement se poursuivra un réseau d'intendants chargé de la fraîcheur des données et de l'identification de nouvelles sources de données sera déployé. La **promotion de la richesse et des usages de scanR sera conduite au sein du Ministère et de ses opérateurs afin que scanR devienne un point de repère pour tous**. scanR est développé et maintenu par le

<sup>10</sup> <https://github.com/dataesr/scanr>

département des outils d'aide à la décision de la sous-direction des systèmes d'information et des études statistiques.

## ➤ **Valoriser et accroître les ressources libres de l'enseignement supérieur – Canal U**

Simplification & Innovation - Action 3 [A03-CONT]

Projet lancé en 2001 par le MESRI avec l'appui de la communauté universitaire et scientifique, **Canal-U est le site de référence pour les ressources audiovisuelles en libre accès de l'enseignement supérieur et de la recherche.** La plateforme offre de multiples possibilités pour valoriser, partager, diffuser et mettre en relation les productions audiovisuelles (cours magistraux, conférences, documentaires...). Le site est opéré et développé par la Fondation Maison des Sciences de l'Homme (FMSH) sous la supervision et avec le financement du MESRI. La crise sanitaire a montré à quel point il était important pour les étudiants d'avoir accès librement à des ressources pédagogiques en ligne. Pour répondre à ce besoin, le Ministère **valorisera les plateformes** comme Canal-U et engagera les établissements d'enseignement à mettre à disposition davantage de contenus accessibles.



Canal-U propose des **ressources pédagogiques et scientifiques à l'attention des enseignants, chercheurs et étudiants.** Elle s'adresse également au **grand public**, pour une formation tout au long de la vie.

### **Canal-U en chiffres (2021) :**

31 500 vidéos  
4 780 000 de visites par an  
Plus de 230 chaînes contributrices (universités, organismes de recherche, musées...)

[www.canal-u.tv](http://www.canal-u.tv)

## ➤ **Enrichir le service de formations en ligne pour les étudiants - FUN**

Simplification & Innovation - Action 4 [A04-CONT]

**FUN**

offre un service de **formation en ligne** et vise à soutenir la formation tout au long de la vie.

### **FUN EN CHIFFRES (2021)**

160 partenaires d'enseignement supérieurs et de recherche  
> 1800 sessions de MOOC diffusées ;  
2,5 millions d'utilisateurs pour 11 millions d'inscriptions à des sessions ;  
800 MOOC dont 25 cours certifiants

<https://www.fun-mooc.fr/fr/>

France Université Numérique opère le service la plateforme FUN et met à disposition des établissements partenaires **des moyens et services mutualisés** pour la diffusion de formation en ligne par les établissements d'enseignement supérieur.

Cette plateforme offre une **visibilité de l'offre française de formations en France et dans le monde.** Son code source est également ouvert<sup>11</sup>.

Le MESRI engage les établissements de l'enseignement supérieur à poursuivre l'enrichissement cette plateforme, pour en faire **l'opérateur national des plateformes de pédagogie.** « FUN ressources » sera créé comme entrepôt national de ressources éducatives libres.

<sup>11</sup> <https://github.com/openfun/>

## ➤ Aider à la création de formations en ligne - *Callisto*

Simplification & Innovation - Action 5 [A05-21S2] puis [A05-CONT]

En avril 2020, le MESRI a demandé aux différents réseaux développant des ressources pour la création de formations en ligne à destination des bibliothèques universitaires de **mutualiser leurs plateformes et leurs ressources**. La **plateforme Callisto** est développée à partir de Moodle, **logiciel libre** d'apprentissage en ligne.

Cette plateforme permettra aussi à de nouveaux partenaires d'intégrer le projet et de créer leurs propres espaces avec leurs propres contenus de formation. Callisto sera lancée en **septembre 2021**.

Plateforme de ressources de formation et d'autoformation à destination des personnels de bibliothèques. Développée sur la base d'un logiciel libre cette plateforme mutualisera les contenus de formations développés par les Centres régionaux de formation aux carrières des bibliothèques (CRFCB) et les Unités régionales de formation à l'information scientifique et technique (URFIST).

Septembre 2021

## Callisto

## ➤ Structurer, ouvrir et partager les données de recherche - *RechercheDataGouv*

Simplification, Innovation & Confiance - Action 6 [A06-22S1] puis [A06-CONT]



recherche.data.gouv.fr

**Recherche Data Gouv** proposera, à partir de mars 2022, une **offre souveraine de dépôt et de signalement des données de recherche** ainsi qu'une offre d'**accompagnement en direction des chercheurs producteurs des données**.

Mars 2022

Pour engager tous les domaines de recherche dans des pratiques actives d'ouverture de données, *Recherche Data Gouv* sera créée. Cette plateforme nationale fédérée des données de la recherche offrira un entrepôt de données pluridisciplinaires qui viendra compléter les infrastructures nationales et européennes spécialisées sur certaines

thématiques disciplinaires. Confié à INRAE et

avec l'appui du CNRS et des universités de Grenoble-Alpes, de Lille, de Lorraine, de Paris, de Paris-Nanterre et de Strasbourg, cet entrepôt au service de la communauté scientifique nationale, proposera un catalogue recensant les données déposées sur d'autres entrepôts de confiance. *Recherche Data Gouv* constituera ainsi à terme le point de visibilité de l'ensemble des données de la recherche française. Fondée sur un **développement et une gouvernance partagés entre les acteurs** de l'enseignement supérieur et de la recherche, *Recherche Data Gouv* garantira à la communauté scientifique la **souveraineté sur les données qu'elle produit**. Les données seront **ouvertes sous licences libres** et la solution de dépôt et de catalogue repose sur un logiciel, dont le **code source est ouvert**. Il dispose également de plusieurs **API ouvertes** pour permettre la recherche, le dépôt et l'accès aux données.

*Recherche Data Gouv* est un **engagement du plan national pour la science ouverte** et est intégralement financé par le Fonds national pour la science ouverte (FNSO) pour ses 3 premières années d'exercice. Cette plateforme déploiera ses premiers services en **mars 2022**.

## ➤ Suivre l'ouverture des données et des codes de la recherche - *Baromètre de la science ouverte*

Confiance - Action 7 [A07-22S2] puis [A07-CONT]



Avec le deuxième plan national pour la science ouverte, le Baromètre de la science ouverte (BSO) sera pérennisé et étoffé. De nouveaux indicateurs seront introduits au-delà des publications pour répondre à différents objectifs :

- **Suivre la dynamique d'ouverture des jeux de données de recherche** associés aux publications et déposés dans un entrepôt disciplinaire ou sur *Recherche Data Gouv*
- **Suivre les déclarations de partage des données** pour celles qui ne peuvent être ouvertes, notamment les données de recherche en santé
- **Suivre dans le temps la production de codes et logiciels** de la recherche française pour en identifier les dynamiques, l'ouverture et les impacts grâce au baromètre de la science ouverte
- **Suivre l'impact des politiques de science ouverte**

Le baromètre de la science ouverte mesure actuellement l'accès ouvert aux publications dans le temps. En 2019 la part des publications mises à disposition librement en ligne concernait 56 % des 156 000 publications scientifiques françaises publiées. L'objectif est de commencer à produire dès 2022 des indicateurs sur la dynamique d'ouverture des données et codes de la recherche.

<https://ministeresuprecherche.github.io/bs/>

Ce baromètre science ouverte des données et des codes sources bénéficie d'un financement du plan de relance sur l'appel ITN5 « cycle de vie de la donnée ».

## B. Gestion des données de la recherche

Amplifier l'ouverture, le partage et l'exploitation des données de la recherche ne pourra se faire sans une gestion maîtrisée des données. Or, la situation en matière de **gestion des données de recherche est à l'heure actuelle relativement critique**. Les chercheurs témoignent des difficultés qu'ils rencontrent : pratiques de stockage sur un support individuel (disque dur, clé USB...), perte des données au moment où un chercheur quitte le laboratoire, absence de solution pour stocker ses données, impossibilité de reproduire les résultats d'une recherche précédente car les données et le code ne sont pas accessibles ou sont mal documentés et non réutilisables.

Les données ouvertes ne constituent qu'une petite frange des données de recherche traitées, notamment parce qu'elles sont très majoritairement associées à des publications, alors que les données « dormantes », difficiles d'accès du fait d'un manque de description, structuration ou préservation dans des environnements communs accessibles, conduisent à une énorme perte de potentiel scientifique et donc à une perte économique évidente.

Toutes les données ne sont pas à conserver. **Le plan de gestion des données devrait constituer un véritable outil de pilotage scientifique** qui permet de définir, en amont d'un projet, les étapes d'acquisition, de nettoyage, de description (curation), de préservation, de destruction et d'ouverture ou partage. Rendre opérant un tel plan permettrait de rendre les données FAIR (Facile à trouver, Accessible, interopérable et réutilisable) dès leur acquisition. L'ambition est de rendre les données de la recherche « FAIR *by design* » pour les doter d'un plus fort potentiel et garantir un stockage maîtrisé et économe des données. Cette feuille de route souhaite porter une **stratégie pour répondre aux difficultés quotidiennes des chercheurs**.

## ➤ Définir une stratégie nationale de services numériques dédiés à la préservation des données de la recherche

Innovation & Simplification - Action 8 [A8-22S2]

La stratégie de transformation numérique, portée par le département Services et infrastructures numériques du MESRI dans le cadre de la labellisation des datacenters sera avantageusement complétée par la définition de services numériques dédiés à la préservation souveraine et sécurisée des données par grands domaines scientifiques. Elle sera ainsi porteuse d'un pas de progrès pour toute la communauté de recherche et d'économies substantielles. Elle permettra de rationaliser les dispositifs et offrira de nouvelles capacités d'analyses ou d'intégration de données non encore valorisées. Cette stratégie suppose un bon dimensionnement des réseaux informatiques sur tout le territoire et proposera, si nécessaire, l'adaptation de l'architecture réseau en place. Le comité Services et Infrastructures Numériques (Co-SIN) dotera l'ESRI d'une stratégie de tels services d'ici fin 2022.

## ➤ Accompagner les chercheurs dans la gestion des données et le « FAIR by design »

Simplification - Action 9 [A9-CONT]

Des « ateliers de la donnée », décrits dans le chapitre IV, seront déployés sur les sites, en proximité des équipes de recherche pour former et accompagner tous les contributeurs de la recherche à la gestion et à la description de leurs données. L'un des enjeux majeurs est de transformer les pratiques pour qu'un plus grand soin soit apporté à la donnée dès son acquisition ou sa production. Le « FAIR by design » qui consiste à rendre une donnée Facile à trouver, Accessible, Interopérable et Réutilisable dès l'amont du cycle de vie de la donnée constituera le levier majeur de cette transformation. Les services d'accompagnement aux chercheurs commenceront à se déployer à partir de début 2022 dans les ateliers de la donnée qui apporteront un appui tout au long du cycle de vie de la donnée.

## C. Préservation et ouverture des algorithmes et des codes sources

Le logiciel joue un rôle clé dans la recherche scientifique, dont il est à la fois un outil, un résultat et un objet d'étude. La mise à disposition des codes sources des logiciels, avec donc la possibilité de les modifier, les réutiliser et les diffuser, est un enjeu majeur pour permettre la **reproductibilité** des résultats scientifiques et soutenir le **partage et la création de connaissances**, dans une logique de science ouverte. Un accès plus libre aux codes sources est également vecteur de **confiance** entre développeurs et utilisateurs.

Afin de mettre en œuvre la politique nationale relative aux données, algorithmes et codes sources<sup>12</sup>, notre ambition est que les codes sources et les logiciels produits par l'administration, l'enseignement supérieur et la recherche publique soient développés et maintenus de façon durable, préservés et valorisés. Dans cette perspective, la fonction d'administrateur des données de la recherche du ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation s'élargit aux algorithmes et aux codes sources.

## ➤ Déposer les codes sources publics du MESRI sur *code.etalab.gouv.fr*

Simplification, Innovation & Confiance - Action 10 [A10-CONT]

Les codes sources développés par l'administration centrale dans le cadre de ses missions sont, sauf cas particulier, indexés sur <https://code.etalab.gouv.fr/>. Cette action sera poursuivie pour permettre

<sup>12</sup> Voir la circulaire du premier Ministre du 27 avril 2021.



aux codes sources publics ouverts produits par l'ensemble des acteurs des services du ministère puissent être à terme répertoriés dans la plateforme [code.etalab.gouv.fr](http://code.etalab.gouv.fr)

## ➤ **Collecter, préserver et partager les codes sources – *Software Heritage*** Simplification, Innovation & Confiance - Action 11 [A11-CONT]

Le MESRI soutient la conservation des codes sources, support indissociable des connaissances techniques et scientifiques. Le deuxième plan national pour la science ouverte poursuit dans cette optique son soutien à **l'archive universelle Software Heritage**. Afin de créer un **écosystème reliant codes, données et publications de la recherche**, la collaboration sera renforcée entre l'archive ouverte nationale HAL (destinée au dépôt et à la diffusion d'articles scientifiques de recherche), la plateforme nationale des données de la recherche *Recherche Data Gouv*, le secteur de l'édition scientifique et Software Heritage.



Software Heritage

Fondée en 2016, *Software Heritage* se charge de **collecter, préserver et partager tous les logiciels disponibles** publiquement sous forme de code source, dans le but de construire une infrastructure commune et partagée au service de l'industrie, de la recherche, de la culture et de la société dans son ensemble.

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

## ➤ **Poursuivre l'ouverture des algorithmes et des données - *Parcoursup*** Confiance - Action 12 [A12-CONT]



La **plateforme Parcoursup** est une démarche simplifiée à destination des **lycéens et étudiants** en recherche d'une réorientation pour s'inscrire dans les formations initiales de l'enseignement supérieur conduisant à des diplômes reconnus par l'Etat.

**Parcoursup en chiffres (2021) :**

- > plus de 19 000 formations, dont plus de 6 000 en apprentissage
- > 931 000 candidats

Le fonctionnement correct des algorithmes de *Parcoursup* est une des pierres angulaires de la confiance que les candidats, leurs familles et les formations placent dans *Parcoursup*.

Dans le cadre des textes en vigueur, le MESRI assure :

- **La publication et la mise à jour régulière du cœur algorithmique** de *Parcoursup*

<https://framagit.org/parcoursup/algorithmes-de-parcoursup>

- **La transparence du processus d'affectation des candidats et l'évaluation de la politique publique d'accès à l'enseignement supérieur** en ouvrant et facilitant l'exploration des données

<https://data.enseignementsup-recherche.gouv.fr/pages/parcoursupdata>

## **D. Usage des logiciels libres**

Les formations dans l'enseignement supérieur reposent sur l'apprentissage et l'utilisation de divers logiciels, sous licences privées ou libres. La question des logiciels s'étend enfin également aux domaines de l'administration et de la gestion dans l'ESR, en englobant les logiciels de scolarité, finances, ressources humaines, etc. Aussi, l'utilisation et la diffusion des productions logicielles comme logiciels libres, c'est-à-dire publiés sous l'une des licences reconnues libres par la Free

Software Foundation et l'Open Source Initiative, seront privilégiées lorsqu'elles seront possibles et pertinentes, comme cela était déjà prévu dans l'article 9 de la loi du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche.

Les logiciels libres sont des leviers **d'innovation** particulièrement forts pour la recherche. Ce point est rappelé dans la stratégie pour le logiciel libre de la Commission européenne de 2020 : la possibilité d'accéder gratuitement aux logiciels, de réutiliser et modifier les codes sources simplifie considérablement la recherche, permettant d'améliorer l'efficacité des chercheurs et des enseignants-chercheurs.

Dans un objectif de **souveraineté**, l'utilisation des logiciels libres vise également à réduire la dépendance de nos systèmes numériques *vis-à-vis* des licences privées ; ils sont en ce sens une bonne réponse aux exigences de la Commission nationale de l'informatique et des libertés (Cnil) formulée en mai 2021 sur la conformité des outils de l'ESR au Règlement général sur la protection des données (RGPD)<sup>13</sup>, préconisant l'usage d'outils alternatifs aux solutions privées. La valorisation des logiciels libres et de l'ouverture des codes sources dans la recherche permet également de se détacher du monde éditeur et ainsi d'assurer une meilleure indépendance des chercheurs.

Plusieurs initiatives ont déjà été mises en place pour développer les usages des logiciels libres dans l'ESR. Concrètement, pour ne citer que celles-ci, des solutions libres comme *Moodle* (environnement de contenus de formation), *Jitsi* ou *BigBlueButton* (logiciel de visioconférence), *LibreOffice* (suite bureautique), *Zimbra* ou *Sympa* (parmi les outils de collaboration) ont déjà été adoptées dans certains établissements de l'enseignement supérieur et de recherche. Les chercheurs et enseignants-chercheurs utilisent aussi déjà de nombreux logiciels libres comme *R* pour le traitement statistique, *Scilab* pour le calcul... Les environnements informatiques de l'ESR reposent aussi très majoritairement sur des solutions libres : *Apache serveur*, *Debian*, *PostgreSQL*...

## ➤ Privilégier l'usage des logiciels libres

Confiance & Innovation - Action 13 [A13-21S2], Action 14 [A14-CONT] & Action 15 [A15-CONT]

Réaffirmer auprès des universités et des organismes de recherche la volonté du Ministère d'adopter le **principe de priorité d'usages des logiciels libres**. Il s'agira d'explicitier ce principe à l'ensemble des opérateurs de l'ESR [Action 13 \[A13-21S2\]](#) et de soutenir, voire généraliser les initiatives prises par différents établissements en matière de conversion vers l'usage des logiciels libres [Action 14 \[A14-CONT\]](#)

Tous les acteurs de l'ESR devront prendre en compte le **Socle interministériel de logiciels libres (SILL)** porté par la **mission dédiée à l'animation et la promotion interministérielles en matière de logiciel libre et de communs numériques** (mission annoncée par la circulaire du Premier ministre en date du 27 avril 2021). Dans le cadre d'un appel d'offres, les établissements de l'ESR devront dûment justifier le recours à un logiciel sous licence privée aux dépens d'un logiciel sous licence. Un cadre opérationnel en ce sens sera fourni par le MESRI aux services achats des opérateurs - [Action 15 \[A15-CONT\]](#)

---

<sup>13</sup><https://www.cnil.fr/fr/la-cnil-appelle-evolutions-dans-utilisation-outils-collaboratifs-etatsuniens-enseignement-superieur-recherche>

## ➤ Encourager la production et l'usage de logiciels libres – *esup portail*

Simplification, Confiance & Innovation - Action 16 [A16-CONT]



Communauté d'établissements de l'enseignement supérieur et de la recherche, *esup portail* déploie une offre de services développée en mode collaboratif par ses adhérents selon 3 principes : **authentification unique, technologies open source et mutualisation**

**esup portail en chiffres (2021) :**

80 établissements adhérents  
11 offres logicielles  
au catalogue

Le MESRI reconnaît et encourage *esup portail* et son action en faveur d'une **offre logicielle souveraine**, basée sur des technologies **open source**, et mutualisée entre les établissements d'enseignement supérieur adhérents.

Le MESRI souligne la simplification engagée au travers de l'authentification unique pour accéder à tous les services. Cette approche pourrait encore se renforcer par la création de partenariats avec la DINUM.

## E. Ecosystème des infrastructures de la science ouverte

Les infrastructures de l'information scientifique sont des infrastructures de la connaissance car elles favorisent la transformation des données en informations, et des informations en connaissances. Ces infrastructures mettent à disposition des outils, des services et des compétences qui répondent aux besoins d'information scientifique tout au long de la chaîne de production des connaissances, c'est-à-dire tout au long du cycle de vie de la recherche, jusqu'aux usagers et publics de celles-ci.

Ces infrastructures assurent ou accompagnent tout ou partie des fonctions de l'information scientifique dans la recherche académique : édition numérique ouverte (*OpenEdition*), édition numérique multi-support structurée et ouverte (*Métopes*), collections et numérisations des grandes bibliothèques scientifiques, collecte et diffusion ouverte des publications scientifiques (*HAL*), préservation et diffusion des codes sources (*Software Heritage*), infrastructure de Text and Data Mining (*ISTEX*), plateforme des vidéos scientifiques (*Canal U*).

Complétées par les entrepôts nationaux disciplinaires de données (*HumaNum* pour les sciences humaines et sociales, *Data Terra* pour les données du système Terre et environnement, *PNDDB* pour les données de biodiversité, *CDS* pour les données d'astronomie...) et de *Recherche Data Gouv* (entrepôt pluridisciplinaire des données de recherche), elles constituent les **infrastructures de la science ouverte**.

## ➤ Développer et animer l'écosystème des infrastructures de la science ouverte

Simplification & Innovation - Action 17 [A17-22S2]

Cet ensemble constitue un paysage dynamique, complémentaire et structuré sans encore former un écosystème au sein duquel seraient développés les liens entre éditions, publications, données et codes logiciels. La première action consisterait à rassembler les porteurs de ces infrastructures en un réseau d'acteurs autour de l'enjeu de **construction collective de l'écosystème français des infrastructures de la science ouverte**. Ces acteurs contribueraient notamment à développer et implémenter la politique des identifiants pérennes et ouverts (cf. Actions 23 à 28).

Il s'agira également de poursuivre la structuration de la communauté française contributrice de l'*European Open Science Cloud (EOSC)*, initiative européenne lancée en 2018 ayant pour objectif de rendre les données scientifiques plus accessibles et de réduire le coût de leur gestion. L'action vise à favoriser activement l'inscription des services français déployés par les infrastructures de la science ouverte dans le catalogue de services de l'*EOSC*.

## F. Simplifier la vie des chercheurs, des enseignants-chercheurs et des étudiants

Dans un souci de mettre en œuvre le « Dites-le nous une fois », le MESRI poursuit son action en direction des usagers de ses services en ligne que sont les étudiants, les enseignants, les enseignants-chercheurs, les chercheurs et les personnels en soutien de l'enseignement supérieur et de la recherche. Pour ces publics spécifiques les actions en direction des chercheurs et enseignants chercheurs seront de natures différentes. Les actions en direction des étudiants se rapprochent de celles développées par d'autres ministères et consistent en la mise en œuvre d'API afin de permettre la circulation des données entre les services de l'état et ainsi éviter de multiples saisies.

Pour les chercheurs, les actions seront de deux natures : faciliter l'accès à des données utiles à leurs recherches et simplifier la soumission de projets dans le cadre des appels d'offre des agences nationales de financement.

Dans le cadre de la politique nationale relative aux données, il s'agira de **favoriser et d'accélérer les demandes d'accès des chercheurs aux données publiques et aux données d'intérêt général.**

### ➤ **Accélérer les demandes d'accès des chercheurs aux données publiques** Simplification - Action 18 [A18-21S2] & Action 19 [A19-22S1]

Pour cela, le MESRI et le MTFP mettent en place au second semestre 2021, une **charte d'engagement de l'Etat Action 18 [A18-21S2]** pour accélérer l'instruction des demandes d'accès aux données publiques (incluant les codes sources et traitements algorithmiques). Cette charte rappelle le cadre juridique applicable et précise les modalités simplifiées de demande d'accès et les délais de réponse sur lesquels s'engage l'Etat. Un *vade-mecum* à destination des chercheurs et des administrations détentrices des données publiques sera édité début 2022 **Action 19 [A19-22S1]**. La mise en œuvre ce dispositif sera évaluée au bout d'un an et révisée, le cas échéant, afin de prendre en compte les retours des chercheurs ou les textes susceptibles d'en impacter le cadre (notamment le *Data Governance Act* européen).

### ➤ **Favoriser l'accès des chercheurs aux données d'intérêt général** Simplification - Action 20 [A20-22S1] & Action 21 [A21-22S2]

Il s'agira de participer aux côtés du MTFP à l'émergence de la **fonction de médiateur des données dites d'intérêt général**, détenues par des **acteurs privés Action 20 [A20-22S1]**. Sur la période que couvre cette feuille de route, l'identification des données d'intérêt général les plus demandées par les chercheurs et enseignants-chercheurs permettra de conduire les premières études de cas. Celles-ci permettront en coordination avec la Direction générale des entreprises de **préfigurer le premier cadre d'indemnisation des détenteurs de données ainsi que les modalités de réutilisation de ces données. Action 21 [A21-22S2]**.

## ➤ Mettre en place des identifiants uniques et pérennes

Simplification - Action 22 [A22-22S1] Action 23 [A23-CONT] Action 24 [A24-22S1]  
Action 25 [A25-21S2] Action 26 [A26-22S1] Action 27 [A27-22S2]

Les chercheurs et enseignants-chercheurs témoignent d'une charge administrative importante liée notamment à la démultiplication de la fourniture de CV, d'identifiants et d'informations diverses pour : leurs candidatures aux appels d'offre, la participation à des missions d'expertise (évaluateurs d'articles soumis à des revues, membres experts de commission d'évaluation, experts dans des instances ou organisations internationales...), l'évaluation par leur établissement employeur ou de leur unité par le HCERES, le dépôt en archives ouvertes de leurs publications, de leurs données et codes sources de recherche. Dans ces différents contextes, les chercheurs doivent gérer de multiples identifiants, et construire des CV en fonction des différents formats imposés.



Identifiant et profil chercheur (CV), reconnu et utilisé internationalement, il permet de se connecter facilement à son écosystème : éditeurs pour soumettre articles..., archives ouvertes de publications, données, voire organisations de recherche

### ORCID en chiffres (2021) :

1000 organisations adhérentes dans le monde  
Consortium ORCID France pour l'adhésion des établissements français

 <https://orcid.org/>

**Poursuivre la promotion de l'adoption de l'identifiant ORCID** – par les chercheurs, afin de consolider leur identité numérique et la visibilité de leurs travaux. Les actions de promotion seront conduites dans tous les établissements. Le consortium ORCID mobilisant l'ABES et Couperin est en appui en portant le site Web en français [Action 22 \[A22-22S1\]](#).

- **Soutenir et participer au pilotage du Consortium Orcid France** qui réunit les établissements adhérents. Ce consortium gère l'adhésion de la France à ce système international indépendant des éditeurs et porte le modèle économique pour la France. Cette action est portée dans le cadre du plan national pour la science ouverte [Action 23\[A23-CONT\]](#).
- **Contribuer au pilotage international** et à l'évolution de ce système d'identifiants afin d'en garantir l'ouverture et l'indépendance sur le long terme [Action 24 \[A24-CONT\]](#).
- Définir et mettre en œuvre une **politique des identifiants uniques des acteurs, structures et objets de la**

**recherche** et soutenir l'alimentation des différents systèmes d'information de la recherche par les données issues d'ORCID pour éviter les doubles saisies. [Action 25 \[A25-22S2\]](#).

Cet identifiant unique et pérenne pour les chercheurs est relié à l'écosystème des identifiants uniques de publications (DOI – *digital object identifier*) et de données (DOI) afin que les résultats de recherche d'une personne lui soient attribués automatiquement sans équivoque.

De façon à ce que cet écosystème garantisse un lien fiable entre les contributeurs de la recherche, leurs productions, leurs laboratoires et leurs établissements :

- Le **répertoire des structures de la recherche française (RNSR)** se transformera pour devenir le **référentiel des structures de la recherche française** et ainsi garantir un point de vérité unique sur lequel les établissements, agences de financement, évaluation... pourront s'appuyer. Le projet de transformation sera cadré début 2022 [Action 26 \[A26-22S1\]](#) et la réalisation sera ensuite conduite dans le calendrier de cette feuille de route. La phase de cadrage fonctionnelle et technique de ce projet définira les modalités de réalisation et de financement de ce projet piloté par le MESRI. **Cette phase de cadrage fera l'objet d'un dépôt de candidature au guichet ITN5 du Plan de relance.**

- La **standardisation du *Software Heritage Identifier (SWHID)*** sera proposée et complètera les identifiants de réalisations de la recherche pour les logiciels aux côtés des DOI des publications et données [Action 27 \[A27-22S2\]](#). Ces travaux seront conduits dans le cadre du plan national pour la science ouverte.

Suivant la démarche « Dites-le nous une fois », ces différentes actions permettront de consolider les **données de références de l'ESRI**, clés de voute pour la **circulation facilitée de données fiables** entre tous les acteurs de l'ESRI français et avec l'écosystème européen et international.

## ➤ **Faciliter les candidatures aux appels à projet de recherche**

Simplification - Action 28 [A28-21S2] Action 29 [29-22S2] & Action 30 [A30-CONT]

Inscrit dans le rapport annexé de la loi de programmation de la recherche (LPR), un portail unique d'accès et de soumission à tous les appels à projets (AAP) portés par des agences nationales de financement sera mis en place, à l'instar de ce qui existe dans d'autres pays. Il sera progressivement enrichi de mesures simplifiant la soumission d'un projet par les chercheurs.

- **1<sup>ère</sup> étape : simplifier l'accès aux informations et disposer d'une vue 360°** sur les appels à projets de 6 agences de financement (ANR, ANSES, INCA, INSERM, ANRS, ADEME). Ce projet coordonné par l'ANR est financé par le MESRI.

**Ouverture du portail d'appels à projet fin septembre 2021** [Action 28 \[A28-21S2\]](#)

- **2<sup>ème</sup> étape : harmoniser les informations et formats des dossiers de candidature entre les agences de financement** et regrouper les appels à projets émanant d'autres financeurs : fondations, régions... Après l'ouverture en septembre 2021 d'une première version, le portail s'enrichira de nouvelles fonctionnalités [Action 29 \[A29-22S2\]](#). Il s'appuiera sur les identifiants uniques ORCID des chercheurs et les identifiants des productions scientifiques pour récupérer automatiquement les résultats de leurs recherches et leur CV. Le financement de cette phase du projet s'appuiera sur une **candidature au guichet ITN2 « transformation numérique des écosystèmes » du plan de relance**, en octobre 2021.

Le portail rassemblera progressivement les appels à projets de différents guichets de financement [Action 30 \[A30-CONT\]](#).


## ➤ **Simplifier les démarches pour les étudiants grâce aux API**

Simplification - Action 31 [A32-21S2 à 22S2]

Les **API statuts étudiants** et **statuts étudiants boursiers** seront déployées afin de simplifier la vie des étudiants. Grâce au développement d'un référentiel national des inscriptions étudiants rassemblant les données d'inscription des plus de 2,7 millions d'étudiants dans plus de 4200 établissements de l'enseignement supérieur, les étudiants n'auront plus à fournir les documents justificatifs de leur statut auprès du CROUS, de la CAF, des collectivités territoriales...qui pourront bénéficier des services de ces API.

## ➤ **Améliorer la plateforme *Trouver mon master***

Simplification & Confiance - Action 32 [A32-23S2]

 le portail national en ligne rassemblant les informations sur les diplômes nationaux de master (capacités d'accueil, modalités d'accès...). Il s'agit de faire évoluer la plateforme pour qu'elle devienne encore davantage le lieu de diffusion de toutes les données relatives aux masters susceptibles d'éclairer les candidats, ainsi qu'une interface unique d'annonce des résultats. La nouvelle plateforme annoncée par la ministre le 8 juillet 2021 donnera **une visibilité complète de l'offre de formations** et permettra une **analyse plus simple et plus rapide**

**des candidatures.** Le système fera en sorte que les délais entre les réponses et les réaffectations de places soient réduits. Ce projet est porté et financé par le DGESIP du MESRI.

## G. Impacts des politiques publiques

### ➤ Engager les établissements de recherche et d'enseignement supérieur à créer et diffuser les jeux de données sur leurs résultats et leurs moyens

Confiance - Action 33 [A33-23S1]

Les établissements de recherche et d'enseignement supérieur publient annuellement leurs rapports d'activité présentant leurs résultats en termes d'atteinte d'objectifs et les moyens consacrés à leurs actions. Pour favoriser une plus grande transparence de leurs politiques, l'agrégation et la réutilisation des données, il leur sera demandé dès fin 2021 d'amplifier ou d'engager une démarche d'open data sur les données. Le MESRI appuiera les établissements en accompagnant la publication des données sur la plateforme *#dataESR*. Cette action devra s'accompagner d'un renfort en compétences du département des outils d'aide à la décision du MESRI. Depuis 2021, les nouveaux contrats passés entre le ministère et ses opérateurs précisent des objectifs en terme d'ouverture des données.

### ➤ Engager les agences de financement à créer et diffuser les jeux de données présentant les projets lauréats et leurs financements

Confiance - Action 34 [A34-CONT]

L'agence nationale de la recherche (ANR), notamment, publie ses jeux de données des **projets lauréats des différents appels à projets**. Ces informations disponibles sur *#dataESR* et *data.gouv.fr* seront progressivement **enrichies par les descriptions des projets et les budgets de financement accordés**. Les autres agences seront invitées à également publier leurs jeux de données. Les données des Programmes d'investissement d'avenir (PIA) viendront également compléter les jeux de données ouverts durant la période de cette feuille de route.

### ➤ Intensifier la transparence des dépenses d'abonnements

Confiance - Action 35 [A35-CONT]

Pour **intensifier la transparence des dépenses** à travers l'ouverture régulière des contrats d'abonnements aux ressources électroniques et l'amplification des collectes de dépenses de frais de publication ("article processing charge"), les informations de dépenses seront publiées annuellement à partir du premier semestre 2022.

### ➤ Améliorer l'usages des services numériques de formation

Confiance - Action 36 [A36-23S2]

La production, la collecte et l'usage de **traces d'apprentissages** permettront de mesurer et adapter les contenus et les usages des formations en ligne. Les expérimentations déjà lancées au sein d'établissements pilotes seront poursuivies afin de favoriser la convergence des données des traces d'apprentissage du monde scolaire et de l'enseignement supérieur dans un entrepôt partagé. Ce projet ambitieux vise à développer et standardiser les mesures d'usage afin d'adapter les offres de formation aux pratiques et attentes des étudiants.

## ➤ **Créer des données sur les usages des infrastructures de la science ouverte**

Confiance - Action 37 [A37-23S2]

La mesure des usages des infrastructures de la science ouverte guidera l'amélioration des services et permettra de qualifier les publics, notamment en identifiant les usages non académiques. Des outils de fréquentation des sites Web des différentes infrastructures sont déjà en place mais ne permettent pas tous une identification précise des usages. L'action consistera à **développer des outils de statistiques de fréquentation normalisées** (par exemple de type COUNTER 5 et notamment *via* Readmetrics) pour toutes les infrastructures de la science ouverte et recommander leur ouverture au public à un grain non identifiant.

## ➤ **Élargir le lectorat des articles scientifiques grâce à la traduction**

Confiance - Action 38 [A38-23S1]

L'objectif est d'élargir le lectorat des articles scientifiques de toutes les disciplines en les sortant du monolinguisme. L'action consistera à réaliser un démonstrateur pour éprouver la faisabilité d'un dispositif de traduction semi-automatique de textes scientifiques. Ce projet s'inscrit dans le cadre d'une politique de science ouverte qui promeut le multilinguisme dans la communication scientifique, avec pour double objectif d'accroître le rayonnement international des recherches françaises publiées en langue native (principalement en sciences humaines et sociales) et d'accroître l'impact sur la société et l'économie françaises des recherches publiées en langue anglaise ou dans d'autres langues étrangères (principalement en sciences, technologies et médecine, mais aussi en économie).

Ce pari de la **traduction profitera aux citoyens et aux décideurs**, qui doivent pouvoir **accéder aux dernières avancées de la recherche** sur la santé, la biodiversité, l'environnement par exemple.

Le démonstrateur visera à tester la faisabilité d'un processus de traduction à grande échelle de textes scientifiques, susceptible de répondre à différents cas d'usage. Il offrira des garanties de qualité grâce à la traduction semi-automatique, associant un moteur de traduction automatique entraîné sur des corpus disciplinaires et une post-édition par des traducteurs professionnels. Il bénéficiera des atouts de la recherche française en matière d'intelligence artificielle et de traitement automatique du langage et s'appuiera sur des briques techniques open source, ouvrant la possibilité d'une réutilisation par les acteurs économiques pour le développement de services numériques innovants.

**Ce projet fera l'objet d'un dépôt de candidature au guichet ITN2 du Plan de relance.**



## IV. Maîtrise des données et des codes sources : acculturation, accompagnement, développement et reconnaissance des compétences

Le développement d'une « culture de la donnée » suffisante est une priorité pour favoriser l'émergence d'une **dynamique de maîtrise et d'utilisation des données** au sein du ministère. Cette acculturation à la gestion maîtrisée des données est aussi une priorité pour les établissements d'enseignement supérieur et de recherche dans lesquels le **soin à apporter aux données à chaque étape du cycle de la recherche doit devenir la pratique**.

Les dispositifs **d'acculturation et de formation doivent être élaborés et se décliner au plus près des enjeux spécifiques de chacun des domaines de données : administration, enseignement supérieur et recherche**.

L'administrateur ministériel des données, des algorithmes et des codes sources est le pilier de la stratégie d'acculturation à la donnée et s'appuiera sur le réseau des services, des établissements d'enseignement supérieur et des organismes pour développer la mise en place de démarches locales, spécifiques et instanciées aux enjeux particuliers.

### A. Actions au sein du MESRI

La **présentation de cette feuille de route constituera la première action phare à l'échelle du MESRI**. Ce **premier événement ouvrira un cycle d'événements récurrents autour de la donnée**. Le ministère dispose d'un capital de données et d'outils de traitement, d'analyse et de diffusion des données encore méconnu d'une grande partie des personnels du ministère et de son écosystème. Avant d'envisager un cycle de formations des personnels, un cycle de présentation des initiatives inscrites dans la feuille de route offrira à chacun l'occasion de découvrir la place des données et des services de valorisation des données déjà déployés au ministère.

De nombreux vecteurs peuvent contribuer à la démarche d'acculturation (présentations en ligne, vidéos pédagogiques en ligne, interventions en présentiel, ateliers spécifiques, documents ou guides de référence, réseaux ou communautés thématiques, etc.). Un équilibre de présentation entre physique et numérique sera à trouver. Les plateformes numériques comme *Canal U* pour les vidéos, *Callisto* pour les formations... seront utilisées pour offrir des présentations et formations à la demande, capitaliser et rediffuser les interventions et formations réalisées en temps réel.

#### ➤ **Mettre en place des « rendez-vous » de la donnée**

Acculturation - Action 39 [A39-CONT]

Un cycle d'événements sera programmé afin de créer les conditions de l'émergence d'une dynamique de maîtrise et d'utilisation des données. En fonction des besoins identifiés avec les managers des services, un programme par alternance des séquences d'informations, d'acculturation aux outils existants, de témoignages, d'ateliers de découverte ou de niveau expert, des formations dédiées à la mise en qualité ou à l'usage des données ou des séquences de mentorat sera défini fin 2021. Les « rendez-vous de la donnée » se dérouleront sur toute la durée de cette feuille de route [Action 39 \[A39-CONT\]](#)

Au-delà de l'émergence de la maîtrise de la donnée et des outils, l'ambition est de développer une culture commune de la donnée au sein du ministère et ainsi créer une émulation collective, vecteur de transversalité et de complémentarité entre services.

## ➤ **Former aux métiers de la donnée**

Action 40 [A40-22S2]

Faire de la donnée un actif stratégique et un bien commun pour tous est un gage d'efficacité pour l'ensemble de l'organisation et de satisfaction pour chacun des acteurs qui y contribue (producteurs des données, intendant en charge de leur mise en qualité et utilisateurs des données en charge de leur analyse et leur valorisation). Mais cela ne sera possible qu'avec un engagement fort de chaque échelon du management. L'évolution des pratiques, les ajustements de l'organisation et la reconnaissance des agents en charge de la collecte, mise en qualité, intendance de la donnée au bénéfice de tous sont des sujets sur lesquels une réflexion collective sera engagée en 2022.

## ➤ **Produire des données de meilleure qualité**

Action 41 [A41-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 des interprétations risquées. Par conséquent, il est nécessaire de s'appuyer sur des données de qualité 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, un groupe de travail sera organisé avec les départements pour définir un réseau d'acteurs en charge de la qualité des données et un processus de gestion de la qualité elle-même adaptée aux usages attendus des données.

Chaque département s'engage à mettre en qualité un à deux jeux de données par an et à déployer le processus de gestion de la qualité en continu sur le ou les jeux nettoyés.

## ➤ **Éditorialiser et rendre intelligible les données pour favoriser leur réutilisation par certains publics**

Innovation - Action 42 [A42-CONT]

Le premier des enjeux de l'ouverture ou du partage des données est la réutilisation par un tiers. Cela nécessite de **diffuser les données sous une forme accessible et intelligible aux autres acteurs** de la société (publics et privés, médias, grand public) que ceux du secteur qui les produisent. Il faut donc que les producteurs des données fassent aussi un **travail de transformation des données en informations qui soient intelligibles et utilisables pour tous les acteurs non spécialistes**. Une bonne transmission à ces acteurs requiert de comprendre les différents besoins et de définir les produits et priorités de transformation. Pour amplifier le **potentiel de réutilisation des données et éviter leur mésusage**, un travail d'édition des données est nécessaire. Il sera conduit en interne du ministère en faisant travailler sur ces questions producteurs et utilisateurs des données.

Chaque département s'engage à rendre intelligible (documenter) un à deux jeux de données par an.

## **B. Accompagner la structuration, le partage et l'ouverture des données et codes sources de la recherche**

Notre ambition est de faire en sorte que les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable), préservées et, quand cela est possible, ouvertes.

La « charge de la donnée » pour les chercheurs est lourde et nous ne pouvons pas nous permettre de laisser les chercheurs seuls face à cette ambition. Pour conseiller **et accompagner concrètement les**

**chercheurs, enseignants-chercheurs tout au long du cycle de vie des données**, la mise en place d'« **ateliers de la donnée** » réunissant une large palette de métiers sera soutenue sur tout le territoire. Ces ateliers seront parties-prenantes de *Recherche Data Gouv* et seront mis en place à partir d'initiatives existantes ou nouvelles. Ils se déploieront dans le cadre des politiques de site pour apporter en **proximité des chercheurs, un premier niveau d'expertise et de services à la gestion, préparation et diffusion des données**. Leurs missions consisteront à :

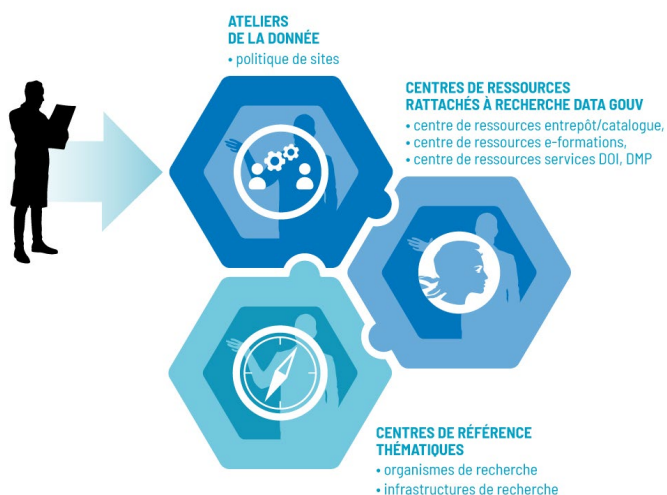
- Accompagner les chercheurs dès l'amont des projets dans la rédaction des plans de gestion de données pour que celui-ci devienne un outil de décision pour la gestion des données tout au long du projet de recherche ;
- Orienter les chercheurs vers les services juridiques ou partenariaux en fonction de la nature des données et du cadre partenarial de production et de traitement des données ;
- Sensibiliser, former et accompagner les producteurs de données à la description des données selon les principes FAIR ;
- Orienter les laboratoires vers les services de stockage, archivage déployés au sein des datacenters ;
- Identifier l'éligibilité des données à l'ouverture ou au partage ;
- Orienter les chercheurs vers l'entrepôt de données de confiance (international ou français) sur lequel déposer leurs données selon le domaine thématique ;

Pour conduire ses missions,

- Les « ateliers de la donnée » s'appuieront sur les expertises scientifiques nationales des centres de référence thématiques en charge de concevoir et porter les référentiels de pratiques thématiques et disciplinaires ;
- Les « ateliers de la donnée » mobiliseront aussi les expertises des services de calcul, stockage du réseau des datacenters ;
- Les « ateliers de la donnée » constitueront et déploieront un réseau de référents de la donnée dans les laboratoires de leur territoire ;
- Les « ateliers de la donnée » seront animés en réseau dans le cadre de *Recherche Data Gouv* afin de mutualiser leurs pratiques, services et formations ;
- Les « ateliers de la donnée » bénéficieront de l'appui d'un centre de ressources qui capitalisera les formations ou services déployés dans des formats de e-formations ou de services en ligne disponibles à la demande.

## ➤ Structurer l'offre des « Ateliers de la donnée »

Action 43 [A43-21S2]



**Les ateliers de la donnée seront le point d'entrée en proximité des chercheurs sur toute nature de besoin relatif à la donnée.** Les initiatives d'accompagnement existantes seront promues et la création de nouvelles initiatives sera encouragée. Cela passera par un appel à manifestation d'intérêt qui sera lancé en octobre 2021. Il s'agira avant tout de mettre en visibilité ces dispositifs au travers du label « Atelier de la donnée ». L'ambition est de développer des synergies cohérentes et complémentaires d'accompagnement sur tout le territoire afin de réduire le « fardeau de la donnée » pour les chercheurs et enseignants-chercheurs.

## ➤ **Soutenir et participer à la *Research Data Alliance* (RDA)**

Action 44 [A44-CONT]

Les pratiques en termes de gestion, de description et d'ouverture des données et codes sources de la recherche se raisonnent et se déploient à l'échelle internationale pour favoriser la circulation et la réutilisation des données et des codes. Au travers du deuxième plan national pour la science ouverte, la France continuera à apporter son soutien à la *Research Data Alliance* (RDA), réseau international définissant les bonnes pratiques dans le domaine des données de la recherche. Les chercheurs et contributeurs de la recherche seront encouragés à poursuivre leur participation aux groupes RDA qui standardisent les pratiques de description des données et des codes sources.

## ➤ **Accompagner l'adoption de licences libres**

Action 45 [A45-CONT]

Afin de favoriser la réutilisation et de protéger la paternité des publications, données et des codes sources de la recherche, l'adoption de licences libres, telles que la licence ouverte (française), une des licences Creative Commons (internationale) ou des licences libres dédiées aux logiciels, doit être favorisée. Des guides d'aide aux choix de licences seront développés et/ou mis à jour par les collègues données et codes sources du comité pour la science ouverte [Action 45 \[A45-CONT\]](#)

## **C. Créer une cohérence entre métiers, formations et parcours**

Produire, préparer, croiser, exploiter, analyser ou visualiser des données nécessite des compétences particulières. De nouveaux types de métiers et de nouveaux profils ont émergé et continue d'émerger. Qu'ils s'agissent d'ingénieurs de la donnée, de *data analysts*, de *data scientists*, *data curator* etc., ces profils sont sous forte tension. Les administrations publiques comme les entreprises peinent aujourd'hui à attirer ces talents car le marché des offres d'emploi dans ce domaine est très concurrentiel. Très conscients de cette priorité, des établissements d'enseignement supérieur ont déjà inscrits ces nouveaux besoins dans leur cursus de formation. Le comité pour la science ouverte a initié des réflexions sur ces questions dans le cadre de son collège des compétences. Pour que la France reste compétitive sur ces métiers, le MESRI en charge de l'enseignement supérieur souhaite porter une politique de formation ambitieuse.

En ce qui concerne le capital humain de l'ESRI, les compétences en matière de gestion, traitement, préservation et ouverture des données et des codes sources des chercheurs, enseignants-chercheurs, personnels d'appui à la recherche et personnels administratifs doivent être développées et reconnues. Cela passera par différentes actions comme le développement de formations continues, les dispositifs d'accompagnement évoqués précédemment, la reconnaissance des contributeurs, l'identification des activités dans les fiches de postes ou la reconnaissance de la diversité des productions de la science ouverte dans les évaluations individuelle et collective.

## ➤ **Introduire une politique de formation initiale à l'usage des données et des codes sources**

Action 46 [A46-CONT]

Les établissements d'enseignement supérieur développent des formations diplômantes des compétences en science et ingénierie de la donnée. Il s'agira de définir une politique de formation aux métiers de la donnée en s'appuyant sur ces initiatives et en les démultipliant. La donnée est présente dans le quotidien de chacun et n'est pas l'apanage de la recherche. Toutes les filières disciplinaires seront encouragées à réfléchir à la mise en place de modules d'acculturation aux usages de la donnée et des codes sources afin de faire des étudiants des citoyens éclairés sur ces questions. Le développement d'unités d'enseignement ou de parcours de formation à l'ouverture et l'usage des données et des codes aux niveaux licence et master viendra renforcer les formations déjà mise en place dans les écoles doctorales.

## ➤ **Reconnaître la diversité des productions et des contributions à la recherche**

Action 47 [A47-CONT]

La mesure 11 du deuxième plan national pour la science ouverte incite à valoriser la science ouverte et la diversité des productions scientifiques dans l'évaluation des chercheurs et enseignants-chercheurs, des projets et des institutions de recherche.

## ➤ **Reconnaître et accompagner l'évolution des carrières**

Action 48 [A48-CONT] Action 49 [A49-CONT] & Action 50 [A50-CONT]

Il s'agira d'accompagner **l'évolution des compétences et des carrières des personnels de la recherche**, notamment en développant des actions de formations continues et en valorisant les compétences et les métiers liés à la gestion du cycle de vie des données et au développement de codes sources :

- Introduire dans le **référentiel de compétences des chercheurs, enseignants-chercheurs, ingénieurs et techniciens de recherche** un sous-ensemble de compétences en lien avec les métiers de la donnée [Action 48 \[A48-CONT\]](#)
- Remettre annuellement un **prix des données de la recherche** pour récompenser les équipes et projets exemplaires en termes de préparation à la réutilisation ou de réutilisation des données de recherche [Action 49 \[A49-CONT\]](#)
- Afin d'accroître la visibilité et la reconnaissance des logiciels comme contribution à la recherche, un **prix du logiciel libre de la recherche** est créé afin de mettre en valeur et récompenser les équipes qui réalisent un travail exemplaire dans ce domaine [Action 50 \[A50-CONT\]](#)

## V. Piloter la feuille de route et évaluer son impact

Le portefeuille de cette feuille de route rassemble 4 familles d'actions :

- Des actions relatives aux données et codes sources de la recherche, inscrites dans le plan national pour la science ouverte. Gérer, préserver, ouvrir ou partager les données et les codes sources sont autant d'actions qui bénéficient aux **progrès de la recherche scientifique, à sa reproductibilité, à l'innovation et à la confiance des citoyens en la science** ;
- Des actions relatives à l'enseignement supérieur qui favorisent **l'accès de chacun aux savoirs et facilitent l'accès des étudiants à un vaste ensemble de ressources pédagogiques** ;
- Des actions de **simplification des démarches à destination des chercheurs et des étudiants** ;
- Des actions au bénéfice d'une **plus grande efficacité collective des acteurs de l'ESRI**.

De nombreux pilotes et parties-prenantes au sein de l'administration centrale et dans les établissements d'enseignement supérieur, de recherche, d'évaluation et de financement sont mobilisés autour de cette feuille de route.

### A. Principes directeurs de la feuille de route

En plus d'adresser une diversité de domaine d'enjeux et d'usages des données et des codes sources, la feuille de route rassemble des actions inscrites dans différents cadres stratégiques : la loi de programmation de la recherche, la stratégie de l'enseignement supérieur, le plan national pour la science ouverte et le plan d'actions de la sous-direction des systèmes d'information et des études statistiques du MESRI en matière d'ouverture des données ou diffusion des études statistiques.

Pour en garantir la réalisation, son ancrage dans l'écosystème des parties-prenantes, son adéquation avec les évolutions numériques et systémiques ainsi qu'avec les attentes des usagers et les usages, cette feuille de route se dote de 3 principes directeurs :

- **Cohérence et programmation globales** : la feuille de route rassemble et programme les initiatives à l'échelle de l'ensemble de l'ESRI en veillant à articuler les projets les uns par rapport aux autres ;
- **Ouverte et à l'écoute de l'écosystème français et international**, la feuille de route adapte sa trajectoire ou crée de nouvelles collaborations en matière de numérique et veille à l'ouverture et l'interopérabilité de ses données, systèmes et services avec l'extérieur ;
- **Usages et usagers** : centré sur les attentes et la participation des usagers dans son élaboration et son évolution, la feuille de route met en œuvre une élaboration collective des services entre porteurs d'idées, de besoins, d'usages et de développements technologiques.

### B. Administrateur ministériel des données, des algorithmes et des codes sources

Le pilotage global de cette feuille de route repose sur l'administrateur ministériel des données, des algorithmes et des codes sources (AMDAC) dont le double défi est d'associer et de coordonner l'ensemble des acteurs mobilisés et créer les conditions favorables d'une meilleure maîtrise, réutilisation et ouverture ou partage des données et des codes sources par la diversité des métiers de l'ESRI.

La fonction d'administrateur ministériel des données, des algorithmes et des codes sources (AMDAC) présente dans chacun des ministères s'incarne au travers de 5 actions clés :

- Élaborer la stratégie du ministère en matière de politique des données, des algorithmes et des sources et la décliner dans une feuille de route pluriannuelle d'actions ;
- Piloter la mise en œuvre de cette stratégie en coordonnant les parties prenantes impliquées dans les projets de la feuille de route ;
- Être le point d'entrée privilégié du ministère sur les sujets de données, algorithmes et codes sources ;
- Organiser l'animation des écosystèmes des producteurs et réutilisateurs des données et codes-sources, afin d'identifier les jeux de données à forte valeur ajoutée, les données manquantes et l'adéquation des services proposés ;
- Conseiller les directions générales en matière de montée en compétences et d'organisation pour faire de la donnée un vecteur de cohérence, d'efficacité et un actif stratégique qui favorise la transversalité des services.

## C. Assurer la mobilisation au sein du ministère et développer les synergies

La mise en œuvre de l'ensemble des actions portées au sein du ministère suppose l'alignement de toutes les parties prenantes impliquées. Cela ne peut se réaliser qu'à condition d'organiser la gouvernance de la donnée collectée, produite, analysée et diffusée au sein de l'ensemble du ministère. **La sous-direction des Systèmes d'Information et des Etudes Statistiques (SIES)** sera en charge de cette mission de gouvernance des données du MESRI puisqu'elle est au centre de l'acquisition et de l'analyse de nombreuses données. Le responsable de cette sous-direction mobilisera les outils et les compétences de celle-ci pour porter cette action ainsi que celles signalées comme lui incombant dans le tableau de bord des actions de cette feuille de route.

Gouverner les données du MESRI pour qu'elles deviennent un bien commun et un actif stratégique favorisera l'efficacité collective mais nécessitera en premier lieu la mise en place d'un catalogue des données.

### ➤ **Cataloguer et décrire les données pour dynamiser les synergies de partage** Gouvernance - Action 51 [A51-23S2]

La mise en place d'une démarche de catalogage de données du MESRI permettra d'établir une vue des données partageables. Les métadonnées qui décriront les jeux de données indiqueront le service producteur des données, la localisation des données, leur signification exacte, les conditions et modalités d'accès, les éléments de protection à prendre en compte, etc. Ces informations stratégiques seront renseignées dans un outil spécialisé en veillant tant que possible à aligner les métadonnées avec celles retenues pour d'autres démarches de catalogage à l'échelle interministérielle afin d'assurer l'interopérabilité des catalogues. Une approche pragmatique sera poursuivie pour augmenter de manière progressive le périmètre des données référencées afin de maîtriser la charge de travail de catalogage et assurer la valeur apportée par le catalogue produit. Cette action sera engagée après appropriation des enjeux et acculturation autour de la donnée de chacun des services et départements. Le catalogue sera maintenu par la sous-direction SIES. L'enjeu de la mise en place et l'actualisation d'un catalogue de données est de désiloter les données collectées et analysées par les services. Cette action permettra concrètement d'initier une description des circuits de données afin de mettre en commun les données au service de tous.

## ➤ **Créer une charte des rôles et responsabilités du capital des données du MESRI** Gouvernance - Action 52 [A52-21S2]

Avant d'initier les travaux de catalogage des données, une charte des rôles et responsabilités des producteurs, intendants et utilisateurs des données sera élaborée et explicitée. Elle définira les rôles de producteur, d'intendant des données en charge de la description et de la mise en qualité des données et d'utilisateur des données. Pour chacun de ces rôles, les responsabilités et compétences seront précisées. Sur la base de cette charte, les différents profils seront identifiés dans les services et/ou département selon un principe de subsidiarité.

## **D. Un réseau d'administrateurs des données, des algorithmes et des codes sources de la recherche**

La constitution du réseau des administrateurs des données, des algorithmes et des codes sources de la recherche des établissements d'enseignement supérieur et de recherche du MESRI est initiée depuis le printemps 2021. Le premier noyau de ce réseau est composé d'une douzaine d'administrateurs placés auprès de la gouvernance de leur établissement. Ce réseau a défini les missions, périmètres d'action, compétences et positionnement des administrateurs des données de la recherche en établissement afin de permettre à chacun des opérateurs du MESRI d'identifier cet acteur clé de la gouvernance des données de la recherche.

## ➤ **Constituer et animer le réseau des administrateurs des données, algorithmes et codes sources de la recherche** Gouvernance - Action 53 [A53-21S2]

Sur la base du cadre de mission défini et du profil décrit, le ministère engagera chacun de ses opérateurs à **nommer un administrateur des données, algorithmes et codes sources de la recherche** en charge de la gouvernance des données et codes sources de la recherche.

## **E. Une gouvernance adaptée au pilotage de la feuille de route**

Cette feuille de route se structure autour de 3 enjeux majeurs communs à la diversité des données et des codes sources portés par le MESRI et à la diversité de ses nombreux opérateurs. Elle fédère des actions pilotées dans des cadres adaptés aux objectifs et aux bénéficiaires de chaque action ou ensemble cohérent d'actions.

Chacune des deux directrices générales président un ou plusieurs comités de pilotage ou comités stratégiques dédiés à des ensembles d'actions. Afin de ne pas démultiplier les instances, l'AMDAC agrégera en un tableau de bord global, les éléments de suivi, indicateurs et priorisations des actions. Ce tableau de bord sera partagé trimestriellement avec les deux directrices générales, le chef des services communs et le Directeur du numérique pour l'éducation afin de permettre le suivi stratégique de la feuille de route. Des arbitrages en termes d'ajustements de la trajectoire, de priorités et de moyens pourront être proposés à ces occasions ou au fil de l'eau en cas de besoin.



## **F. Mesure des impacts des actions de la feuille de route**

Les impacts de la feuille de route seront mesurés en rassemblant les indicateurs définis à l'échelle de chaque action ou ensemble d'actions. Cette feuille de route compte parmi ses actions un certain nombre de dispositifs d'évaluation : suivi de l'ouverture des données et des codes, traces d'apprentissage, usages des infrastructures de la science ouverte... Un cadre cohérent d'indicateurs d'adoption des services par les usagers sera proposé à l'échelle de la feuille de route. Plus globalement les impacts tangibles dans la transformation publique et l'évolution des pratiques d'enseignement et de recherche seront appréciés à mi-parcours et à la fin de cette feuille de route.

## Actions faisant l'objet de dépôts de candidature au plan de relance



### ➤ Mettre en place des identifiants uniques et pérennes Simplification Action 27 [A27-22S1]

Le *répertoire des structures de la recherche française (RNSR)* se transformera pour devenir le **référentiel des structures de la recherche française** et ainsi garantir un point de vérité unique sur lequel les établissements, agences de financement, évaluation... pourront s'appuyer. Le projet de transformation sera cadré début 2022 [Action 27 \[A27-22S1\]](#) et la réalisation sera ensuite conduite dans le calendrier de cette feuille de route. La phase de cadrage fonctionnelle et technique de ce projet définira les modalités de réalisation et de financement de ce projet piloté par le MESRI. **Cette phase de cadrage fera l'objet d'un dépôt de candidature au guichet ITN5 du Plan de relance.**

### ➤ Elargir le lectorat des articles scientifiques grâce à la traduction Confiance - Action 38 [A38-23S1]

L'objectif est d'élargir le lectorat des articles scientifiques de toutes les disciplines en les sortant du monolinguisme. L'action consistera à réaliser un démonstrateur pour éprouver la faisabilité d'un dispositif de traduction semi-automatique de textes scientifiques. Ce projet s'inscrit dans le cadre d'une politique de science ouverte qui promeut le multilinguisme dans la communication scientifique, avec pour double objectif d'accroître le rayonnement international des recherches françaises publiées en langue native (principalement en sciences humaines et sociales) et d'accroître l'impact sur la société et l'économie françaises des recherches publiées en langue anglaise ou dans d'autres langues étrangères (principalement en sciences, technologies et médecine, mais aussi en économie).

Ce pari de la **traduction profitera aux citoyens et aux décideurs**, qui doivent pouvoir **accéder aux dernières avancées de la recherche** sur la santé, la biodiversité, l'environnement par exemple.

**Ce projet fera l'objet d'un dépôt de candidature au guichet ITN2 du Plan de relance.**

### ➤ Faciliter les candidatures aux appels à projet de recherche Simplification - Action 28 [28-21S2]

Un portail unique d'accès et de soumission à tous les appels à projets (AAP) de la recherche française sera mis en place, à l'instar de ce qui existe dans d'autres pays.

- **1<sup>ère</sup> étape : simplifier l'accès aux informations et disposer d'une vue 360°** sur les appels à projets de 6 agences de financement (ANR, ANSES, INCA, INSERM, ANRS, ADEME). Ce projet coordonné par l'ANR est financé par le MESRI.  
**Ouverture du portail d'appels à projet fin septembre 2021**
- **2<sup>ème</sup> étape : harmoniser les informations et formats des dossiers de candidature entre les agences de financement** et regrouper les appels à projets émanant d'autres financeurs : fondations, régions... Après l'ouverture en septembre 2021 d'une première version, le portail s'enrichira de nouvelles fonctionnalités [Action 29 \[A29-22S2\]](#). Il s'appuiera sur les identifiants uniques ORCID des chercheurs et les identifiants des productions scientifiques pour récupérer automatiquement les résultats de leurs recherches et leur CV. Le financement de cette phase du projet s'appuiera sur une **candidature au guichet ITN2 « transformation numérique des écosystèmes » du plan de relance**, en octobre 2021.





**MINISTÈRE  
DE L'ENSEIGNEMENT  
SUPÉRIEUR,  
DE LA RECHERCHE  
ET DE L'INNOVATION**

*Liberté  
Égalité  
Fraternité*

