























## CHAPITRE 1 – JE DOCUMENTE MES DONNEES

Documenter ses données est une partie très importante dans l'activité de recherche. Il s'agit d'expliquer :

- comment elles sont créées, numérisées,
- ce qu'elles contiennent (dictionnaire de variables...),
- leur structure (la structure logique qui permet d'organiser et traiter les données),
- et également l'ensemble des manipulations dont elles ont fait l'objet.

Cette étape répond à des enjeux de confiance dans la **qualité des données** pour l'accès, la mise à disposition, la réutilisation et l'archivage pérenne.

### 1.1. AVOIR RECOURS AUX METADONNEES

Une bonne documentation des données inclut nécessairement d'avoir recours à des métadonnées. Les métadonnées sont un ensemble de données structurées décrivant des ressources physiques (documents papier tels que rapports ou ouvrages...) ou numériques (fichier PDF d'articles, tables excel, photographies...). Ces informations sont souvent structurées selon des schémas internationaux (DUBLIN CORE, ISO 19115 pour les informations géographiques...).

**Ces métadonnées sont un maillon essentiel pour le partage de l'information et l'interopérabilité des bases de données et des systèmes informatiques associés.** Cette étape est cruciale et indispensable pour la mise à disposition et/ou l'archivage et détermine le cycle de vie de la donnée. Des données insuffisamment documentées ne peuvent pas être réutilisables, ni archivées. Ces éléments de description sont les garants d'une production de données fiables et authentiques.



## 1.2. LE CATALOGAGE DES DONNEES IRSTEADATA

Les objectifs du catalogue sont les suivants :

- Offrir un catalogue institutionnel des données en vue de préparer l'Institut aux exigences Inspire,
- Rendre accessible en interne l'ensemble des jeux de données dont nous disposons.

Pour ce faire, nous disposons d'un catalogue interne permettant de partager des jeux de données disponibles et réutilisables au sein d'Irstea (produites mais aussi acquises).

Il est accessible à cette adresse : <http://data.Irstea.fr>

### 1.3. ENREGISTRER AU BON FORMAT

Le format de fichier dans lequel vous conservez vos données est un facteur déterminant de la manière dont vos données seront conservées et réutilisées dans le futur.

Comme les évolutions technologiques sont constantes, il est absolument nécessaire de planifier la gestion de l'obsolescence des logiciels ainsi que du matériel informatique permettant de lire et d'exécuter les programmes à l'origine de la création des données. Comment vos données seront-elles accessibles et lues si le logiciel utilisé pour les produire devient inopérant ?

Il est communément admis que les formats les plus susceptibles d'être accessibles à l'avenir sont ceux qui répondent aux critères suivants :

- non propriétaire,
- ouvert,
- documenté,
- dont l'usage est reconnu par la communauté scientifique concernée,
- répondant à des standards (ASCII, Unicode),
- non crypté et non compressé.

Il est donc fortement recommandé de migrer vos données dans un format qui répond aux caractéristiques ci-dessus, et de conserver en plus une copie dans le format original du logiciel.

Pour les images, se référer au guide d'utilisation de la photothèque accessible sur l'intranet.

## ENCADRE : DONNEES GEOGRAPHIQUES

La Directive du 14 mars 2007, dite Directive « Inspire », vise à fixer les règles générales destinées à établir l'infrastructure d'information géographique dans la Communauté européenne (ci-après dénommé « Inspire »), aux fins des politiques environnementales communautaires et des politiques ou des activités de la Communauté susceptibles d'avoir une incidence sur l'environnement" (article 1).

La Directive définit les termes suivants :

- ✚ « Donnée géographique », toute donnée faisant directement ou indirectement référence à un lieu ou une zone géographique spécifique ;
- ✚ « Série de données géographiques », une compilation identifiable de données géographiques.

La directive s'applique aux séries de données géographiques qui sont en format électronique (les données sur support papier ne sont pas concernées) et qui sont détenues par "une autorité publique, après avoir été produites ou reçues par une autorité publique, ou bien gérées ou mises à jour par cette autorité et rentrant dans le champ de ses missions publiques".

Sont considérés comme autorité publique :

- (a) tout gouvernement ou toute autre administration publique, y compris les organes consultatifs publics, au niveau national, régional ou local ;
- (b) toute personne physique ou morale exerçant, dans le cadre du droit national, des fonctions dans l'administration publique comportant des tâches, des activités ou des services en rapport avec l'environnement ;
- c) toute personne physique ou morale ayant des responsabilités ou des fonctions publiques, ou fournissant des services publics en rapport avec l'environnement, sous le contrôle d'un organisme ou d'une personne visé au point a) ou b).

Les données concernées sont classées dans des thèmes figurant à l'annexe I, II ou III. Seules les données concernées par les thèmes des annexes sont des données Inspire.

Ces annexes énumèrent 34 thèmes qui couvrent l'essentiel des données de référence et de nombreuses données thématiques environnementales ou en rapport avec l'environnement (santé, population, bâti, réseaux, agriculture etc.). Ces séries de données sont soumises à des règles spécifiques de diffusion (services en réseau) et de partage entre administrations (mutualisation complète).

## CHAPITRE 2 – JE SECURISE ET PERENNISE MES DONNEES

### 2.1. EVALUER SON BESOIN DE SECURITE

Avant d'engager une démarche technique de sécurisation de ses données, il est pratique de connaître et d'évaluer son besoin de sécurité pour les données. En effet, toutes les données manipulées n'ont pas besoin du même niveau d'exigence en matière de sécurité.

#### Qu'est-ce qu'une donnée sensible ?

En droit français, une donnée sensible est une donnée à caractère personnel qui peut permettre d'établir un lien avec :

- des origines raciales ou ethniques,
- des opinions politiques, philosophiques ou religieuses,
- une activité syndicale,
- la santé des personnes,
- les orientations sexuelles.

Il est, a priori, interdit de collecter et d'enregistrer de telles données.

### 2.2. METTRE EN ŒUVRE LES ACTIONS DE SECURITE

La mise en œuvre des actions de sécurité est variable suivant le niveau de sécurité identifié :

- Le besoin de sécurité « minimum » est couvert par les actions de maintien en condition opérationnelle de la Direction des Systèmes d'Information (DSIN).
- Le besoin de sécurité « standard » nécessite de faire une analyse de risque complète pour laquelle vous devez vous adresser au Responsable Sécurité des Systèmes d'Information (RSSI) qui vous accompagnera pour cette analyse
- Les actions permettant de satisfaire le besoin de sécurité « élevé » doivent être mises en place aussi avec l'accompagnement du RSSI.

#### 2.2.1. QUELLES SONT LES REGLES A SUIVRE ?

Il sera toujours utile pour vous de vous adresser à votre Responsable des Systèmes d'Information (RSI) ou Administrateur des Systèmes d'Information (ASI) pour la prise en charge de votre besoin en lui indiquant les différents éléments de sécurité que vous avez identifiés ainsi que le niveau global de votre besoin.

---

## 2.2.2. SECURITE TECHNIQUE

---

### 2.2.2.1. FIXER DES REGLES DE STOCKAGE

Le stockage, notamment sur des moyens centraux, est le moyen de faire prendre en charge par les équipes informatiques l'hébergement de vos données. Cela permet en outre d'assurer la disponibilité et l'intégrité des données (notamment avec les sauvegardes).

Des règles simples de stockage et de partage de fichiers permettront d'assurer l'unicité matérielle et temporelle (intégrité) des données, mais aussi des programmes et procédures.

Il est conseillé de stocker les données brutes et le traitement qui a permis d'obtenir les résultats (reproductibilité) avec les modèles utilisés.

---

### 2.2.2.2. PREVOIR CAPACITE ET COUT DE STOCKAGE

En amont et dès le début de votre projet, il est essentiel de traiter l'aspect technique avec votre RSI ou votre ASI dans votre centre pour que ce dernier puisse à la fois déterminer la volumétrie de votre besoin au départ et au cours de votre projet et identifier les financements nécessaires.

## 2.3. ARCHIVER LES DONNEES

L'archivage des données est une étape incontournable de tout projet de recherche. L'objectif principal de la fonction d'archivage est de choisir, conserver et de rendre disponible les documents ou informations qui ont une valeur patrimoniale. Les archives de données ne font pas exception, mais elles diffèrent de certaines archives notamment en matière d'infrastructures techniques à mettre en œuvre compte tenu de la masse de données produite de façon exponentielle depuis quelques années.

L'archivage correspond à des enjeux de **pérennisation, d'accessibilité et de traçabilité des données**. Cet enjeu doit être pris en compte à toutes les étapes du cycle de vie de la donnée.



## CHAPITRE 3 – JE M’ASSURE DES DROITS DE PROPRIÉTÉ DE MES DONNÉES ET DES CONDITIONS D’UTILISATION

### 3.1. LA PROTECTION JURIDIQUE D’UNE DONNÉE

La donnée brute, telle que définie page 8, n’est pas protégée juridiquement.

Toutefois, le fait de réunir et de structurer les données entre elles est protégeable juridiquement.

La protection d’une base de données est régie par le code de la propriété intellectuelle. Elle peut se faire par deux moyens, indépendants l’un de l’autre (art L.341-1 du CPI) :

- par le droit d’auteur, qui régit la protection de l’auteur, ce dernier étant considéré comme celui qui a fait le choix, de manière originale, de la disposition et des matières utilisées pour la réalisation de la base,
- par le droit sui generis, qui régit la protection du producteur, c’est-à-dire de celui qui a réalisé les investissements financiers, matériels ou humains substantiels nécessaires à la création de la base. A ce titre, le producteur est généralement l’employeur.

L’auteur et le producteur sont, sauf cession expresse, titulaires de la base.

Le titulaire d’une base est en droit de décider de son utilisation/exploitation. Cependant, il peut être soumis à certaines restrictions légales ou contractuelles sur sa base de données.

Pour en savoir plus sur la protection des bases de données (BDD) :

La **donnée personnelle** se distingue des autres données dans la mesure où elle nécessite une attention particulière de son titulaire.

Elle est définie par l’article 2 de la loi 78-17 du 6 janvier 1978 modifiée dite « Informatique et libertés » comme « toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement » (ex : adresse mail, n° de sécurité sociale etc.). Pour traiter ou diffuser la donnée d’une personne physique, il est obligatoire d’obtenir son consentement préalable. Lors de la diffusion, le consentement de la personne concernée sera nécessaire pour chaque usage que l’on souhaite faire de la donnée.

Tel qu’il est mentionné au point 2.1 de ce guide, il peut également être interdit de collecter certaines données personnelles. C’est le cas lorsqu’elles sont considérées comme « sensibles » (article 8 de la loi dite « Informatique et Libertés »), il s’agit notamment de données relatives à la santé, la politique ou la biométrie.

### **La diffusion de la donnée personnelle est libre si elle a été anonymisée :**

En effet, Le **10 avril 2014**, le G29 réunissant les « CNIL » européennes a rendu un avis relatif aux techniques d'anonymisation des données à caractère personnel. Dans son avis, le G29 rappelle que toute technique d'anonymisation doit être construite au cas par cas et adaptée aux types de traitements de données. Afin d'évaluer l'efficacité d'une technique d'anonymisation, le G29 préconise de procéder en trois étapes. Il vérifie (i) qu'il est impossible d'isoler un individu à partir de l'ensemble de données, (ii) qu'il est impossible de relier entre eux des ensembles de données distincts concernant un même individu, (iii) qu'il est impossible de procéder par déduction pour obtenir de l'information sur un individu. Si ces trois critères sont remplis, le G29 considère que l'ensemble de données est bien anonyme. Si au moins un des critères n'est pas rempli, une analyse détaillée des risques de ré-identification sera nécessaire pour juger de l'efficacité de la technique d'anonymisation. (Suite sur Europa.eu)

## **3.2. LES DROITS ET OBLIGATIONS SUR UNE BASE DE DONNEES**

Par principe, le titulaire d'une base de données est libre de décider de son utilisation (l'exploiter via des licences, la mettre en libre accès, la céder etc.) mais il peut être soumis à certaines restrictions légales ou contractuelles auxquelles il se doit d'être attentif.

### **Le cadre légal :**

D'une part, l'accès à certaines données peut ou doit être refusé lorsqu'il s'agit :

- de données à caractère personnel
- de données portant atteinte au droit de propriété intellectuelle
- de données portant atteinte aux secrets protégés par la loi
- Autres cas de refus potentiel : il est possible de refuser l'accès à une donnée pour protéger l'environnement auquel ces informations ont trait, tels que les sites de reproduction d'espèces rares.

D'autre part, l'Europe et la France font état d'une volonté générale, au travers de la directive INSPIRE, de la convention Aarhus, à ce que les bases de données produites par les autorités publiques soient au moins accessibles librement par tout public.

Un accès libre n'est pas pour autant synonyme d'une utilisation libre. La réutilisation de la base de données peut être limitée par une licence ou une convention de réutilisation et être soumise à redevance afin que le titulaire puisse percevoir un retour financier sur l'investissement consacré à cette base de données.

### **Le cadre contractuel :**

Par ailleurs, hormis les obligations légales, le titulaire de la base de données devra également s'assurer du droit des tiers sur cette base. Notamment, il lui appartient de vérifier qu'il n'est pas en copropriété sur cette base (la base de données est-elle le fruit d'un partenariat ?), qu'il n'a pas concédé de licence exclusive à un tiers ou s'il n'a pas cédé sa base.

### **3.3. LES DIFFERENTS TYPES DE CONTRAT D'EXPLOITATION D'UNE BASE DE DONNEES**

Sous réserve du droit des tiers et du cadre légal mentionné au point 3.2 de ce guide, le titulaire d'une base de données peut choisir de :

- céder sa base de données, gratuitement ou non,
- licencier sa base de données, à des fins d'utilisation et/ou d'exploitation, gratuitement ou non.

L'intérêt de donner libre accès à ces données est multiple :

- une plus grande notoriété de l'EPST et des recherches menées,
- une meilleure diffusion des connaissances, notamment entre les autorités publiques,
- faciliter la réalisation d'études et ainsi le suivi et l'évaluation de politique menée (politique environnementale, par exemple).

Néanmoins, cela nécessite de prendre quelques précautions (propriété, responsabilité etc.) et d'encadrer juridiquement leur réutilisation.

### **3.4. LA DEFENSE DE LA BASE DE DONNEES**

Les bases de données juridiquement protégées ne peuvent pas être utilisées sans l'autorisation du titulaire de la base, sous peine d'être condamné pénalement. Réciproquement, il est possible d'agir contre un tiers utilisant frauduleusement notre base de données.

## **CHAPITRE 4 – JE VALORISE MES DONNEES**

### **4.1. QU'EST-CE QUE LA VALORISATION DES DONNEES ?**

La valorisation des données est une activité qui tend à considérer les données comme une ressource économique, patrimoniale et technique. Cette valorisation peut être de deux ordres :

- économique : considérer les données comme une ressource permettant de tirer un revenu suite à son exploitation (§ 4.2.1),
- non économique : considérer les données comme une ressource patrimoniale et/ou technique dont il faut assurer l'identification, l'accès et la mise à disposition (§ 4.2.2).

Les activités de valorisation des données recouvrent plusieurs aspects : d'ordre technique (rendre accessible par les métadonnées – cf. § 1.1., rendre exploitable par des formats ouverts, mettre en place des systèmes d'information interopérables – cf. § 2.), d'ordre juridique (assurer la protection des producteurs et la réutilisation par la mise en place de licences – cf. § 3.), d'ordre économique (licencier des bases de données, les vendre ou vendre des services autour de la base de données,



fixer un prix – cf. § 4.2.1), d'ordre marketing (connaissances des besoins des utilisateurs, communication sur l'existence des bases – cf. § 4.2.1).

## 4.2. LES DIFFERENTS MODELES DE VALORISATION

- ✚ Diffusion libre et gratuite des données brutes (exemple : data.eaufrance)
- ✚ Diffusion gratuite des données brutes, paiement pour les données pour lesquels un travail d'enrichissement particulier a été réalisé
- ✚ Diffusion payante des données brutes ou des données travaillées
- ✚ Redevances pour des services autour de la diffusion des données (service d'extraction à la demande, prestation de conseil, etc.)
- ✚ Valorisation des logiciels permettant d'exploiter les données

#### 4.2.1. VALORISATION ECONOMIQUE

Penser à la valorisation de ses données, c'est se poser la question de l'origine des données et de leur documentation. Autrement dit, une valorisation sans risque passe nécessairement par les étapes indiquées aux § 1 et 2 et s'envisage dès la construction du projet scientifique dans le cadre duquel les données seront produites.

On rencontre trois situations :

1. Mon unité de recherche rattachée à IRSTEA collecte elle-même les données, sans acquérir de données auprès de tiers, les données ne sont pas des données personnelles ou ne sont pas considérées comme sensibles ou relevant du secret médical. A partir de ces données, l'unité de recherche crée une base de données en interne.

→ Il n'y a pas d'entrave à une valorisation des données ainsi que de la base.

2. Les données qu'utilise mon unité de recherche sont acquises auprès de tiers (par exemple, un autre laboratoire de recherche ne dépendant pas d'IRSTEA ou un prestataire). Il faut se poser les questions suivantes :

- Ai-je acquis les droits de propriété intellectuelle (cela peut aussi se poser dans le cadre d'un marché public) ?
- Dans quelles limites puis-je les exploiter ? (y a-t-il des conditions associées à l'usage des données ?)

→ Dans ce cas, il peut y avoir des entraves pour réexploiter les données pour une autre recherche ou dans un cadre de valorisation. Il faut vérifier en amont les conditions associées et les respecter.

3. Le programme de recherche s'inscrit dans le cadre d'un partenariat scientifique, d'un programme européen ou international. Le contrat de partenariat règle-t-il la propriété :

- Des données d'origine ?
- Des données résultant du processus de recherche ?

→ Il faudra étudier et respecter les termes du partenariat scientifique ou du programme européen.

Cas des bases de données développées avec des partenaires : Si la base de données est en copropriété avec d'autres partenaires, il faut rédiger un accord de copropriété. Pour cela, rapprochez-vous de la DIVAC pour négocier au mieux les conditions de la copropriété. Une fois la fiche de négociation remplie, la DIVAC envoie le document à la DIVAC qui rédige un accord de copropriété.

Cas des bases de données développées par des UMR : Il faut lire les clauses de propriété intellectuelle dans les conventions d'UMR. Si les bases de données appartiennent en copropriété aux tutelles du laboratoire, les décisions relatives à la valorisation sont prises en commun et la signature par toutes les tutelles des accords relatifs à la valorisation est souvent nécessaire. En revanche, en général, une seule tutelle est en charge de réaliser la valorisation.



---

## COMMENT PROTÉGER MA BASE DE DONNÉES ?

Les moyens de protection juridique d'une base de données sont détaillés dans le chapitre 3.2 et dans la fiche pratique sur le site web sur les données <https://donnees-recherche.irstea.fr/>

Pour matérialiser la protection d'une base de données, il est conseillé de faire un dépôt à l'Agence de Protection des Programmes (APP).

---

## COMMENT FAIRE CONNAÎTRE MA BASE DE DONNÉES AUPRES DES UTILISATEURS ?

Une fiche d'offre technologique peut être rédigée afin de promouvoir votre base de données. Une offre technologique comprend notamment une description détaillée de la base de données, une liste des avantages apportés, les applications potentielles, le stade de développement et les responsables à contacter.

La DIVAC rédige l'offre technologique en lien avec l'équipe de recherche.

La fiche sera diffusée par la DIVAC sur différents portails et bourses de technologies.

---

## QUELS SONT LES MODÈLES ÉCONOMIQUES DE VALORISATION ?

Pour en savoir plus sur les différents modèles économiques de valorisation : consul

---

#### 4.2.2. OPEN ACCESS ET OPEN DATA : VALORISATION PAR LA LIBRE MISE A DISPOSITION

Ces deux mouvements se développent en parallèle et tendent à s'interconnecter depuis quelques années grâce notamment aux moyens technologiques de plus en plus puissants. Mais ils relèvent de deux philosophies dont les racines sont différentes, l'un est né de la volonté de transparence et de responsabilité des autorités publiques vis-à-vis des citoyens, l'autre est au départ un mouvement de chercheurs engagés pour une plus grande accessibilité de leurs résultats scientifiques pour un accès libre au savoir.

Le libre accès (en anglais : *open access*) est la mise à disposition en ligne de contenus numériques, qui peuvent eux-mêmes être soit libres (Creative commons, etc.), soit sous un des régimes de propriété intellectuelle. Une des premières déclarations internationales majeures sur le libre accès, qui inclut une définition, une information de fond et une liste de signataires, est l'*Open Access Initiative* de Budapest de 2001. Ce rassemblement est reconnu comme le premier rassemblement historique fondateur du mouvement libre accès.

---

#### PERIMETRE DE L'OUVERTURE DES DONNEES PUBLIQUES

Concernant l'ouverture des données publiques, il faut distinguer deux aspects :

- **L'accès aux données**

La réglementation française qui s'applique en la matière est la loi de 1978 régissant l'accès aux documents administratifs. La notion de documents administratifs est entendue de façon très large. Il existe des restrictions à l'accès aux données notamment toutes les données ou documents relatifs à l'exécution d'un contrat de prestation de service, les documents relevant de la défense nationale, du secret médical ou du secret industriel et commercial. Toute personne peut obtenir l'accès à des documents et données dans ce cadre auprès de l'administration sur simple demande en se tournant vers la CADA (<http://www.cada.fr/>) en cas de refus ou d'absence de réponse de la part de l'administration.

- **La réutilisation des données**

La réutilisation s'entend comme l'utilisation par toute personne à d'autres fins que celles de la mission de service public dans laquelle les données ont été produites. Les restrictions à la réutilisation sont liées au caractère industriel et commercial des données et au droit de propriété intellectuelle d'un tiers.

La directive 2003/98/CE du 17 novembre 2003<sup>2</sup> concernant la réutilisation des informations du secteur public, transposée en droit français en 2005, fixe un ensemble minimal de règles concernant la réutilisation et les moyens pratiques destinés à faciliter cette réutilisation.

---

<sup>2</sup> (<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:FR:PDF>)



Ces « documents » peuvent être interprétés de façon large et prendre la forme de données. La directive 2013/37/UE du Parlement européen et du Conseil du 26 juin 2013 modifiant la directive de 2003 se traduit par la mise en place d'un principe de réutilisation des données publiques. Cependant les données produites par les organismes de recherche demeurent exclues du champ d'application de la directive. La transposition de la directive ne devrait pas bouleverser la législation en la matière dans la mesure où la réutilisation de toutes les données publiques quelques soient leurs origines sont par principe réutilisables sauf contre-indications liées au droit des tiers ou à des exclusions contractuelles. En l'état actuel de la réglementation, les organismes de recherche peuvent fixer eux-mêmes les conditions de la réutilisation de leurs données (ordonnance N°2005-605 du 6 juin 2005).

En résumé, les obligations de l'administration afin de garantir une réutilisation sans risque sont de s'assurer que les données ne puissent pas être altérées et dénaturées, de mentionner la source et d'indiquer la date de la dernière mise à jour.

L'administration peut déterminer un prix / redevance pour la réutilisation en fonction de trois éléments : le coût de la collecte et de la production, le coût de traitement et le coût de mise à disposition de ses données. La directive de 2013 impose que le prix appliqué pour la réutilisation des données ne dépasse pas le coût marginal de la production de la donnée. Cependant, la tendance actuelle et générale en Europe est de privilégier la gratuité de la réutilisation notamment en raison de la complexité des redevances et du peu de gains perçus (rapport Trojette)

Il existe de nombreuses licences de réutilisation (la licence ouverte d'Etalab, la licence ODBL de l'Open Knowledge Foundation, la licence de l'APIE). Il est important de bien savoir à quoi autorise la licence rattachée aux données qu'on réutilise car ces licences engagent le réutilisateur et parfois de manière contradictoire entre elles.



---

## DONNEES ET PUBLICATIONS SCIENTIFIQUES

Il existe plusieurs modèles de publication des données liées aux articles :

- Les données sont publiées sous forme de fichiers supplémentaires associés aux articles,
- Les données sont déposées dans des bases de dépôts - soit dans des formats répondant à des normes et standards internationaux reconnus, soit sous forme de fichiers génériques avec un certain niveau de métadonnées, et reliés aux articles de revues des éditeurs,
- Les données sont publiées sous la forme de flux structuré et lisible par une machine (XML par exemple).

Ces modèles sont en évolution et il existe aujourd'hui de nombreux défis techniques pour relier de façon opérationnelle et pérenne les articles et données à leurs auteurs.

Pour les données diffusables en libre accès, Irstea a déjà organisé la diffusion sur Internet, d'une partie des publications de ses auteurs (versions auteur, scientifiquement validées des post-publications et dans le respect des embargos imposés par les éditeurs) au travers de son archive institutionnelle, Irstea Publications-CemOA, à partir de laquelle un reversement partiel s'effectue dans l'archive nationale HAL. Le choix Irstea du dépôt obligatoire des publications de ses auteurs depuis 1992 est aujourd'hui l'un des moteurs de sa politique (cf. référence de la note).

Avec ce dispositif, Irstea répond déjà aux nouvelles obligations contractuelles mises en place par l'UE avec Horizon 2020 ; Irstea fait en outre évoluer son archive institutionnelle de publications pour la rendre compatible avec OpenAire, suivant le même modèle que pour HAL, afin de maintenir un dépôt unique et obligatoire pour les auteurs Irstea tout en garantissant un accès le plus large possible aux publications qui peuvent ou doivent être diffusées en libre accès.

De même, la mise en place d'une Photothèque scientifique et technique permet à chaque équipe Irstea de déposer ses photos et bientôt des vidéos réalisées dans le cadre des travaux de recherche. Il s'agit là aussi d'un dispositif de collecte et de diffusion à la fois interne ou externe.

Irstea a par ailleurs mis en place un catalogue des jeux de données et bases de données un outil de scientifiques de l'Institut IrsteaData. Le libre accès à une partie de ce catalogue est prévu.

A ce jour, aucun dispositif n'est en place pour garantir la diffusion des données amont aux publications. Or, H2020 rend obligatoire la diffusion de certaines données collectées ou produites sur ce type de financement. Les équipes scientifiques devront donc analyser, avec leurs partenaires, et pour chaque projet, les moyens nécessaires à la structuration et à la diffusion des données collectées, acquises et/ou produites dans le cadre des travaux de recherche. L'appel à un prestataire externe (institutionnel ou privé) est envisageable si aucune solution interne satisfaisante n'est disponible au démarrage du projet n'est disponible.

## CITER LES DONNEES POUR UNE MEILLEURE VISIBILITE

La démarche de diffuser et/ou de publier ses données permet de les rendre visibles, et d'augmenter par ce biais la visibilité du chercheur, celle de ses publications et de son laboratoire.

Aujourd'hui la meilleure façon de rendre visibles les données, si les conditions juridiques et de réutilisation, sont compatibles, est l'identifiant numérique d'objet, en anglais le « Digital Object Identifier » (DOI). Il s'agit d'un identifiant unique et pérenne rattaché à un document, fichier, donnée...toute ressource numérique.

## COMMENT JUGER LA CREDIBILITE DES BASES DE DEPOT DES DONNEES ?

Il est important de s'assurer d'un certain nombre d'informations avant de déposer ses données dans des bases ouvertes et de vérifier au moins trois éléments :

- **Qui ?** Il est important de regarder qui est à l'initiative de la base : universités (ex : Dryad) ou éditeurs de revue (base « ScientificData » de Nature <http://www.nature.com/scientificdata/>).
- **Pourquoi ?** Identifier l'objectif de la base : partenariats avec d'autres institutions et/ou éditeurs, interconnexion avec d'autres bases, la politique des conditions de réutilisation des données, les services proposés, le système de tarification...
- **Comment ?** Le processus de dépôt est-il clairement décrit et la protection juridique (les licences / droits d'auteur / copyright) associée à la base de dépôt ainsi que les conditions d'utilisation des données que vous souhaitez déposer sont-elles lisibles ?

## QUI CONTACTER AU SEIN D'IRSTEA ?

Pour toute question relative aux thématiques abordées dans le présent guide, vous pouvez contacter l'équipe données-R2 via ce formulaire : <https://donnees-recherche.irstea.fr/contact/>

## AUTEURS DU GUIDE

Contributeurs : Caroline Martin (DRISE), Aurélie Charpentier (DIVAC), Laurence Freyt-Caffin (DIVAC), Nicolas Raidelet (DSIN), Fabrice Moutte (DSIN), Céline Delmas (DIVAC), Maud Pilon (DQ), Emmanuelle Jannes-Ober (DRISE), animateur du groupe de travail et chef de projet : Soizic Rugani (DQ). Mise à jour 2017 : Colette Cadiou ( DRISE), Emmanuelle Jannes-Ober ( DRISE)

Relecteurs : Patrice Meriaux (OHAX), Eric Quinton (EABX), Dominique Laborde (DSRE), APIE (Agence du Patrimoine Immatériel de l'État).

Participants au brainstorming initial : Nicolas Raidelet (DSIN), Daniel Uny (ADBX), Fabrice Moutte (RSSI), Laurence Freyt-Caffin (DIVAC), Nadia Guiffant (MECF), Maud Pilon (Déléguée Qualité processus recherche), Sonia Grimbuhler (TSAN), Aurélie Charpentier (DIVAC), Eric Maldonado (RSI Grenoble), Caroline Martin (DRISE), Flora Branger (HHLY), Soizic Rugani (Déléguée Qualité unités support), Joëlle Franquet (Assistante Qualité), Patrice Meriaux (OHAX), Pierre Maurel (TETIS), Guy Roussin (TETIS).

## LISTE DES FICHES PRATIQUES DISPONIBLES

Les fiches pratiques sont regroupées sous 5 rubriques sur le site <https://donnees-recherche.irstea.fr/>

- Gérer et documenter
- Protéger, utiliser : sécurité et cadre juridique
- Valoriser
- Partager, publier
- Réutiliser

## GLOSSAIRE

### **Archivage** :

Ensemble de pratiques et d'outils utilisés pour conserver des informations stables (mortes) et pouvoir les consulter quels que soient leurs dates, format et support matériel.

Fonction ayant pour objectif la fixation et la préservation de la preuve sur toute la période nécessaire. Cette fonction intègre l'organisation et la gestion d'un contrôle efficace et systématique de la création, de la réception, de la conservation, de l'utilisation et du sort final des documents et des informations.

*Outils de management – Management de l'information – FD X50-185 : 2004-03*

### **Base de données** :

Recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique, et individuellement accessibles par des moyens électroniques ou par tout autre moyen.

*Code de la propriété intellectuelle, art. L112-3 - [http://www.legifrance.gouv.fr/citoyen/new\\_code.ow](http://www.legifrance.gouv.fr/citoyen/new_code.ow)*

Ensemble de données organisées, en vue de son utilisation, par des programmes correspondant à des applications distinctes et de manière à faciliter l'évolution indépendante des données et des programmes.

*Glossaire informatique des termes de la Commission ministérielle de terminologie informatique - <http://www.rocq.inria.fr/qui/Philippe.Deschamp/CMTI/glossaire.html>*

En informatique, système de gestion d'ensembles de données (fichiers) et de leurs relations.

*Outils de management – Management de l'information – FD X50-185 : 2004-03*

**Catalogage** : action qui permet d'informer sur les données disponibles et sur les modalités pratiques d'accès à ces données.

### **Donnée** :

En informatique, on appelle donnée la représentation conventionnelle, après codage, d'une information sous une forme permettant d'en faire le traitement électronique.

*Dictionnaire encyclopédique de l'information et de la documentation / Serge CACALY (dir.)- Nathan, Paris : 1997.- p. 59*

Représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement. Arrêté du 22/12/1981.

*CRITER (base de données terminologiques de la délégation générale à la langue française qui comprend l'ensemble des termes (près de 3000) publiés au Journal officiel par la Commission générale de terminologie et de néologie) - <http://www.culture.fr/culture/dglf/terminologie/base-donnees.html>*



Par "donnée", on pourrait entendre, au sens étroit du terme, une information formatée pour être traitée par un système informatique. Elle sera entendue ici au sens large d'information collectée ou produite sur n'importe quel support, pas seulement informatique.

*Circulaire du 14 février 1994 relative à la diffusion des données publiques, dite "circulaire BALLADUR", Journal Officiel de la République Française : 19 février 1994 - [http://www.legifrance.gouv.fr/citoyen/jorf\\_nor.ow?numjo=PRMG9400081C](http://www.legifrance.gouv.fr/citoyen/jorf_nor.ow?numjo=PRMG9400081C)*

La donnée est un élément immatériel correspondant à un fait, un état ou un événement du réel, par lequel l'être humain appréhende son environnement. Elle peut avoir plusieurs formes : textuelle, graphique, sonore, voire "informatique" c'est-à-dire, transcodée en binaire sur des supports magnétiques, optiques...

*SANDRE, Approche de la donnée version 1.0 / Ministère de l'environnement, 1995*

### **Données brutes** :

Jeu de données initial provenant d'une source extérieure constituant la matière première sur laquelle vont s'appliquer les différents traitements. Ce jeu de données est souvent accompagné de documents descriptifs, clauses d'utilisation et de dictionnaires qui serviront à enrichir le jeu de métadonnées associé. Ces données doivent être archivées et ne doivent subir aucune modification.

Matière élémentaire de la connaissance

Représentation réinterprétable d'une [d'un élément] d'information, sous une forme conventionnelle convenant à la communication, à l'interprétation ou au traitement [XP X 50-435]

*Outils de management – Management de l'information – FD X50-185 : 2004-03*

**Données géographiques** : Les données géographiques sont des données « faisant directement ou indirectement référence à un lieu ou une zone géographique spécifique ».

*Directive 2007/2/CE du 14 mars 2007 établissant une infrastructure d'informations géographiques dans la Communauté Européenne (INSPIRE)*

**Données délivrables** : Jeu de données issu d'un ou de plusieurs traitements effectués en suivant un cahier des charges plus ou moins formel à partir de données brutes. Ce fichier est accompagné de métadonnées. Ce jeu de données est plus connu des chercheurs sous le nom de "fichier(s) de travail" puisque le processus de recherche empirique est réalisé à partir de ces fichiers exploitables. Ces fichiers peuvent être reproduits à l'identique ou modifiés en suivant la chaîne des traitements décrits dans la démarche qualité.

**Intégrité** : Unicité matérielle et temporelle des fichiers, des procédures et des programmes

### **Métadonnée** :

Une métadonnée (mot à mot "donnée à propos de donnée") est une donnée servant à définir ou décrire le contenu d'une autre donnée quel que soit son support. Elle peut être utilisée pour décrire une donnée sous la forme de catalogue, ou expliciter plus précisément le contenu d'un fichier ou d'une table. Elle permet en outre, de partager et de publier des informations dans un format commun.

Métadonnées : Ensemble structuré de données servant à localiser et à décrire une ressource informationnelle consignée sur un support documentaire, en vue de faciliter et d'améliorer son repérage, sa gestion, son usage ou sa préservation.

*Outils de management – Management de l'information – FD X50-185 : 2004-03*

### **Sauvegarde** :

Opération qui permet de répondre au risque de perte ou de corruption de données (effacement par erreur, corruption par bogues, pannes matérielles, catastrophes). Une opération de sauvegarde réalise une copie d'un ensemble cohérent de données telles qu'elles existaient à un moment bien identifié dans le temps. Les objectifs des opérations de sauvegarde sont la restauration des données,



c'est-à-dire la persistance des données même après un incident majeur. La sauvegarde ne suffit donc pas, il faut pouvoir, après une perte, restaurer une image cohérente des données.

Action qui consiste à garantir la préservation de tous documents ou informations par des moyens adaptés eu égard aux risques identifiés.

*Outils de management – Management de l'information – FD X50-185 : 2004-03*

## POUR EN SAVOIR PLUS

Normes de référence : FDX 50-550, FDX 50-551 (2001) en cours de mise à jour (NF X 50-553 : management des activités de recherche) et norme expertise NFX 50-110. Normes de catalogage ISO 19115 et 19139.

## BIBLIOGRAPHIE

### Documentation générale

- OCDE « Principes et lignes directrices pour l'accès aux données de recherche financée sur fonds publics » <http://www.oecd.org/fr/sti/sci-tech/38500823.pdf>
- Déclaration de Berlin <http://openaccess.mpg.de/286432/Berlin-Declaration> (juillet 2003)
- International Council for Science Union (ICSU) : <http://www.icsu.org/> et <http://www.icsu-wds.org/organization/data-policy>
- Recommandations de la Commission européenne 17 juillet 2012 relative à l'accès aux informations scientifiques et à leur conservation : [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/recommendation-access-and-preservation-scientific-information\\_fr.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_fr.pdf)

### Accessibilité et partage des données : enjeux et perspectives

- Socha, YM (ed.) CODATA-ICSTI Task Group on Data Citation Standards and Practices (2013) Out of cite, out of mind: the current state of practice and technology for the citation of data: Data Science Journal, Vol. 12 <http://dx.doi.org/10.2481/dsj.OSOM13-043>
- Fecher B., Friesike S, Hebing M., What drives academic Data sharing” Working paper Series of the german Data Forum (RatSWD) avril 2014 [http://www.ratswd.de/dl/RatSWD\\_WP\\_236.pdf](http://www.ratswd.de/dl/RatSWD_WP_236.pdf)
- Berman F. & Cerf, V. (2013). Who Will Pay for Public Access to Research Data? Science, 341(6146), 616-617 <http://dx.doi.org/10.1126/science.1241625>
- Jing-Woei Li Open Data: A Sustainable Model Science 29 November 2013: Vol. 342 no. 6162 pp. 1042 DOI:10.1126/science.342.6162.1042-a
- Costas, R., Meijer, I., Zahedi, Z. & Wouters, P. (2013). The Value of Research Data - Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report [www.knowledge-exchange.info/datametrics](http://www.knowledge-exchange.info/datametrics)
- Piwowar, H. (2013). Altmetrics: Value all research products. Nature, 493(7431), 159–159 <http://dx.doi.org/10.1038/493159a>
- Piwowar, H.A. & Vision, T.J. (2013). Data reuse and the open data citation advantage. PeerJ PrePrints 1:e1v1 <http://dx.doi.org/10.7287/peerj.preprints.1v1>
- Report on Knowledge Exchange workshop: Price of keeping knowledge: financial streams for digital preservation. Amsterdam (2013) <http://www.knowledge-exchange.info/Default.aspx?ID=570>

- Bellini, E. u. a. (2012). APARSEN Persistent Identifiers Interoperability Framework, Public Report. Fondazione Rinascimento Digitale, Florence, Italy  
[http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D22\\_1-01-1\\_9.pdf](http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D22_1-01-1_9.pdf)
- Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Suhonen, J., Wilson, M., Science, H., et al. (2012). Compilation of results on drivers and barriers and new opportunities.  
<http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/08/ODE-CompilationResultsDriversBarriersNewOpportunities1.pdf>
- GRDI2020. (2012). Global Research Data Infrastructures: The Big Data Challenges. GRDI2020 Final Roadmap Report. <http://www.grdi2020.eu/Repository/FileScaricati/e2b03611-e58f-4242-946a-5b21f17d2947.pdf>
- Kowalczyk, S. & Shankar, K. (2013). Data Sharing in the Sciences. Annual Review of Information Science and Technology, 45(1), 247–294.  
<http://dx.doi.org/10.1002/aris.2011.1440450113>  
<http://onlinelibrary.wiley.com/doi/10.1002/aris.2011.1440450113/pdf>
- Jahnke, L. & Asher, A. (2012). The Problem of Data. Council on Library and Information Resources, Washington, D.C. ISBN 978-1-932326-42-0  
<http://www.clir.org/pubs/reports/pub154> <http://www.clir.org/pubs/reports/pub154/pub154.pdf>
- The Royal Society (2012). Science as an open enterprise. The Royal Society Science Policy Centre. ISBN:978-0-85403-962-  
[http://royalsociety.org/uploadedFiles/Royal\\_Society\\_Content/policy/projects/sape/2012-06-20-SAOE.pdf](http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf)
- Thomson Reuters (2012). Collaborative Science: solving the issues of discovery, attribution and measurement in data sharing.  
[http://wokinfo.com/products\\_tools/multidisciplinary/dci/collaborative\\_science\\_essay.pdf](http://wokinfo.com/products_tools/multidisciplinary/dci/collaborative_science_essay.pdf)
- Uhler, P.F. (2012). For Attribution — Developing Data Attribution and Citation Practices and Standards. Summary of an International Workshop. National Research ISBN: 9780309267281 [http://www.nap.edu/catalog.php?record\\_id=13564](http://www.nap.edu/catalog.php?record_id=13564)
- Wouters, P., Beaulieu, A., Scharnhorst, A. & Wyatt, S. (eds) (2012). Virtual Knowledge. The MIT Press. ISBN: 9780262018395 <http://mitpress.mit.edu/books/virtual-knowledge-0>
- Van den Eynden, V., Corti, L., Wollard, M., Bishop, L. & Horton, L. (2011), Managing and Sharing Data: a best practice guide for researchers, 3rd ed., University of Essex, Colchester, Essex, UK. <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>
- Van der Graaf, M. & Waaijers, L. (2011). A Surfboard for Riding the Wave. Towards a four country action programme on research data. A Knowledge Exchange Report.  
<http://www.knowledge-exchange.info/surfboard>



## Les plans de management des données

- European Commission « Guidelines on Data Management in Horizon 2020 »  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- Des sites de référence:  
<http://www.nerc.ac.uk/research/sites/data/dmp.asp?cookieConsent=A>  
<http://www.dcc.ac.uk/resources/data-management-plans>

cours en ligne

- [http://datalib.edina.ac.uk/xerte/play.php?template\\_id=2](http://datalib.edina.ac.uk/xerte/play.php?template_id=2)  
<http://datalib.edina.ac.uk/mantra/libtraining.html#profiles>
- Les guides en ligne du JISC (UK) :  
<http://www.jisc.ac.uk/guides/research-data-management>
- Les guides en ligne du MIT (USA) :  
<http://libraries.mit.edu/guides/subjects/data-management/>
- Les guides en ligne de la bibliothèque de l'université de Californie :  
<http://www.cdlib.org/services/uc3/datamanagement/>
- Les guides en ligne de la bibliothèque de l'université du Minnesota :  
<https://www.lib.umn.edu/datamanagement>

## Les politiques des plans de management de données

- CDL list US Funder Policies on their DMPTool  
site [https://dmp.cdlib.org/pages/funder\\_requirements](https://dmp.cdlib.org/pages/funder_requirements)
- DMPs in the Australian Code for the Responsible Conduct of Research  
<http://ands.org.au/guides/data-management-planning-awareness.html> (voir aussi <http://monash.edu/library/researchdata/guidelines/data-planning/>)
- DMP encouraged : <http://www.dans.knaw.nl/en/content/data-management-plan>  
and <http://www.dans.knaw.nl/en/content/data-archive/data-contracts>
- Recommandation pour l'exigence d'un DMP dans les projets et propositions financés dans le cadre de H2020: [http://recodeproject.eu/wp-content/uploads/2013/04/2013\\_OARD\\_Recommendations\\_public.pdf](http://recodeproject.eu/wp-content/uploads/2013/04/2013_OARD_Recommendations_public.pdf)
- En Allemagne : [http://www.allianzinitiative.de/en/core\\_activities/research\\_data/principles/](http://www.allianzinitiative.de/en/core_activities/research_data/principles/)
- La directive 2003/98/CE du 17 novembre 2003 (<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:FR:PDF>)

### Archivage des données

- Archivage électronique Un nouveau domaine d'expertise au service de la gouvernance des systèmes d'information  
[http://references.modernisation.gouv.fr/sites/default/files/DISIC\\_AE\\_Guide\\_bonnes\\_pratiques\\_0.pdf](http://references.modernisation.gouv.fr/sites/default/files/DISIC_AE_Guide_bonnes_pratiques_0.pdf)

### L'usage et droit des données et des bases de données

- Guibault, Lucie; Wiebe, Andreas (Eds) (2013) SAFE TO BE OPEN: STUDY ON THE PROTECTION OF RESEARCH DATA AND RECOMMENDATION FOR ACCESS AND USAGE) <http://webdoc.sub.gwdg.de/univerlag/2013/legalstudy.pdf>

### Nouveaux processus de publication et « data journals »

- Callaghan, S., Murphy, F., Tedds, J., Allan, R., Kunze, J., Lawrence, R., Mayernik, M.S. & Whyte, A. (2013). Processes and Procedures for Data Publication: A Case Study in the Geosciences. International Journal of Digital Curation, 8(1), 193–203.  
<http://dx.doi.org/doi:10.2218/ijdc.v8i1.253>
- Priem, J. (2013). Scholarship: Beyond the paper, Nature, 495(7442), 437– 440  
<http://dx.doi.org/10.1038/495437a>
- Grootveld, M. & van Egmond, J. (2012). Peer-Reviewed Open Research Data: Results of a Pilot. International Journal of Digital Curation, 7(2), 81–91  
<http://www.ijdc.net/index.php/ijdc/article/viewFile/221/290>  
<http://dx.doi.org/10.2218/ijdc.v7i2.231>

### Open data et données personnelles

- Rapport d'information sur l'open data et laprotection de la vie privée, Gaëtan Gorce et François Pillet, Sénat, avril 2014.  
<http://www.senat.fr/rap/r13-469/r13-469.html>
- La protection des informations à caractère personnel dans le cadre de l'ouverture et du partage des données publiques Mémento COEPIA, juillet 2013  
[http://www.gouvernement.fr/sites/default/files/fichiers\\_joints/coepia\\_memento\\_donnees\\_personnelles.pdf](http://www.gouvernement.fr/sites/default/files/fichiers_joints/coepia_memento_donnees_personnelles.pdf)
- Avis n°12 du Conseil national du numérique relatif à l'ouverture des données publiques («Open data»), CNN, 5 juin 2012  
[http://www.cnnnumerique.fr/wp-content/uploads/2012/06/2012-06-05\\_AvisCNNum\\_12\\_OpenData.pdf](http://www.cnnnumerique.fr/wp-content/uploads/2012/06/2012-06-05_AvisCNNum_12_OpenData.pdf)

<sup>1</sup> *Recommandation du Conseil concernant l'accès aux données de recherche financée sur fonds publics, OCDE, 14 décembre 2006*



---

<sup>i</sup> Les décrets d'application de la LRN sur le TDM ne sont pas encore promulgués (attendus pour mai 2017)