



Fiche pratique n°3 : Bonnes pratiques pour la gestion des informations dématérialisées de la recherche scientifique

Fiche pratique

Bonnes pratiques pour la gestion des informations dématérialisées de la recherche scientifique

Projets de recherche, études, bases de données scientifiques permanentes

Intérêt des informations et disponibilité dans le temps

Au sein des organismes de recherche, les **informations recueillies et produites** dans le cadre de la recherche et les bases de données scientifiques constituent leur **cœur de métier**. De plus en plus, ces données existent **uniquement sur support électronique** et pour certaines, comme les bases de données, ne peuvent pas être matérialisées sans perdre leur intérêt fonctionnel.

Leur conservation électronique est donc **primordiale** :

- sur le moyen terme : répondre aux besoins fonctionnels des producteurs (accéder à l'information contenue) ; obligations d'ordre légal, réglementaire, stratégique, contractuel, exploitation commerciale des résultats (brevets, licences, participation).
- à long terme : dans une perspective historique et patrimoniale, dans une perspective de sauvegarde du savoir scientifique pour réutilisation éventuelle de données brutes ou pour la référence de sources pour des résultats obtenus.

Pour les bases de données contenant des données identifiantes, une gestion rigoureuse est nécessaire pour répondre aux exigences imposées par la **CNIL**.

Spécificités des informations dématérialisées

Quand le support de conservation des informations est le papier, l'information et son support sont indissociables. Concernant les fichiers électroniques, l'information n'est accessible qu'au moyen d'un dispositif de lecture complexe impliquant du matériel et du logiciel. Or, cet appareillage technologique est particulièrement fragile et peut rendre les données inaccessibles ou très chères à récupérer si l'on constate :

AURORE, groupe de travail sur les archives scientifiques, fiche pratique n°3

- L'illisibilité des supports d'enregistrement : chute d'un disque dur, contrôle irrégulier ou stockage inadapté d'une collection de CD, de DVD...
- L'enregistrement des informations dans un format propriétaire non publié, un logiciel de lecture qui ne fonctionne plus sous un nouveau système d'exploitation, la disparition d'un éditeur de logiciel.

Actuellement, la durée de vie des technologies numériques varie de cinq à dix ans, tandis que les besoins de conservation s'étendent bien au-delà de ces limites.

Un autre phénomène lié à la dématérialisation des pratiques de gestion informationnelle est **l'éparpillement potentiel des informations** appartenant au même sujet à des endroits d'enregistrement différentes (**boîtes courriels, serveurs, dossiers sur disque dur, etc.**) avec des identifications qui ne sont pas nécessairement codifiées et rendent la réunification postérieure des informations difficile.

Cette fiche vous propose ...

... un guide succinct des bonnes pratiques de gestion de vos informations en suivant les étapes de production des données électroniques dans un contexte de recherche scientifique. Sous cette forme, il vous accompagnera tout au long de votre projet.

Ces bonnes pratiques s'appliquent autant à la production d'information issue de bases de données et plateformes de recherche qu'aux données produites pendant des projets de recherche.

Les différentes bonnes pratiques sont présentées d'abord sous forme de MEMENTO. Leur développement IN EXTENSO se trouve à la suite.

Pourtant, même la partie IN EXTENSO n'est qu'une introduction générale aux problématiques et solutions envisageables. La pratique de la gestion et l'adaptation de ces bonnes pratiques à votre cas reste à développer par vos soins.

N'oubliez surtout pas que vous n'êtes pas seuls dans la mise en place de la bonne gestion de vos informations : vos interlocuteurs experts – archivistes et informaticiens – sont à votre service.

!!!!!!

**Sollicitez les archivistes et informaticiens
dès la phase de préfiguration de vos projets scientifiques
pour mettre en place une gestion adéquate de vos informations
et garantir la disponibilité de vos données
tout au long et au-delà de votre recherche**

!!!!!!

Bonnes pratiques de la gestion des informations dans le contexte de recherche scientifique (projets de recherche, base de données, ...)

MEMENTO

Phase de préfiguration du projet

- ✓ Commencer à établir et recueillir la documentation de la base de données ou du projet de recherche
- ✓ Prendre contact avec le service d'archives et le service informatique de l'organisme
- ✓ Prendre en compte les documents internes et externes déjà existants qui imposent des règles de gestion de l'information
- ✓ Rédiger un plan d'assurance qualité (PAQ) de la gestion des informations du projet qui définit les règles et rôles pour la gestion des informations (en prenant en compte tous les points suivants) → *conseils pour classement, nommage et tri à la fin du document*

Phase de production

- ✓ Tous les contributeurs doivent être informés des règles de gestion des informations
- ✓ Désigner un nombre limité de contributeurs qui contrôleront la gestion documentaire du dossier
- ✓ Disposer d'un endroit unique d'enregistrement des informations
- ✓ S'imposer des règles de nommage communes des fichiers
- ✓ S'imposer des règles de classement des fichiers
- ✓ Intégrer la correspondance avec l'extérieur
- ✓ Assurer un lien clair entre les parties dématérialisées et les parties matérialisées (papier)
- ✓ Séparer la documentation scientifique extérieure de la production propre
- ✓ Appliquer un choix restreint et pas trop « exotique » des formats des fichiers et données

Phase de finalisation

- ✓ Vérification par le chef du projet de l'intégralité des données et informations dans le dossier/endroit unique d'enregistrement puis clôture officielle
- ✓ Appliquer les modalités de transmission pour archivage, les critères de tri et les durées de conservation

Bonnes pratiques de la gestion des informations dans le contexte de recherche scientifique (projets de recherche, base de données, ...) IN EXTENSO

Phase de préfiguration du projet

- ✓ **Etablir, dès la phase de préparation, la documentation de la base de données ou du projet de recherche** qui présente finalité, outils, moyens, méthodes et ressources mis en œuvre, le programme et la planification : protocole de l'étude, questionnaire ou instrument de recueil des données, dossier CNIL/CCTP/CCTIRS/convention(s) de partenariat, résultats/rapports de l'étude etc.
Pour l'archivage, ces documents ont un intérêt particulier : ils fournissent des métadonnées sur le projet de recherche ou la base de données, la documentation permettant de **comprendre tout le contexte** de production des données. Il est donc important d'archiver également ces documents.
D'autres informations et documents décrivant des **aspects plus techniques** peuvent s'y rajouter. Notamment pour les bases de données, les documents suivants sont essentiels pour comprendre le fonctionnement de l'outil :
 - **Dictionnaires des champs et des données** : tableau avec nom (codifié et en clair) des champs, longueur/taille du champ (nombre de caractères), type de donnée/valeur (numérique, texte etc.) par champ, format des valeurs par champ (longueur fixe ou variable, formatages spécifiques comme date etc.), liste des variables possibles (nom et codification) pour les champs à choix de valeurs restreint
 - **Structure des enregistrements (=dessin du fichier)** : composition des enregistrements (comme export brut des valeurs saisies) champ par champ dans l'ordre d'apparition en indiquant les positions des valeurs correspondants aux champs (position de départ et d'arrivée dans l'enregistrement par addition successive des nombres de caractères)
 - **Structure de la base s'il s'agit d'une base relationnelle** : présentation (souvent figurative ou en tableau) des tables d'enregistrement (nom de la table, indication du nombre et des noms des champs contenus dans la table) et des relations entre les tables d'enregistrement (clés primaires etc.)

Afin de faciliter l'archivage, les métadonnées peuvent être collectées dès la création du document et tout au long de son cycle de vie. Cela participe à une gestion documentaire rigoureuse et de qualité.

- ✓ **Prendre contact avec le service d'archives et le service informatique de l'organisme avant le lancement du projet** et les associer à la rédaction ou la relecture avant validation des documents de préfiguration. Ces interlocuteurs vérifieront la faisabilité au niveau de la disponibilité des informations du projet dans le temps et mettront en place, le cas échéant, les outils nécessaires pour y parvenir.
- ✓ **Prendre en compte avant le lancement des documents internes et externes déjà existants qui imposent des règles de gestion de l'information** : stratégie de la gestion de l'information de l'organisme, documents Qualité, plans de classement, règles de gestion documentaire, tableau de gestion des archives, obligations légales ou contractuelles, etc.
- ✓ **Rédiger un plan d'assurance qualité (PAQ) de la gestion des informations du projet qui définit les règles et rôles pour la gestion des informations** (en prenant en compte tous les points suivants), notamment la répartition des rôles, tâches et responsabilités à l'intérieur de l'équipe, les droits d'accès et d'utilisation, la liste des documents et informations à déposer obligatoirement, plan de nommage, structure de classement, etc.

Phase de production

- ✓ **Tous les contributeurs doivent être au courant des règles de gestion des informations**
- ✓ **Désigner un nombre limité de contributeurs qui contrôleront la gestion documentaire du dossier** : ils auront la responsabilité et la tâche de vérifier l'application par tous des règles de gestion communes, de faire les corrections et de contrôler que l'intégralité de toutes les informations nécessaires soient enregistrées.
- ✓ **Disposer d'un endroit unique d'enregistrement des informations**, accessible par tous les contributeurs selon leurs droits d'accès
- ✓ **S'imposer des règles de nommage communes des fichiers** : le nom d'un fichier est adéquat lorsque le contenu du document est compréhensible grâce au titre. Il s'agit d'établir un plan de nommage avec règles générales et, pour certains documents obligatoires et indispensables, de définir au préalable leurs titres. → *Voir quelques conseils de nommage en fin de ce document.*
- ✓ **S'imposer des règles de classement des fichiers** : définir une arborescence des répertoires électroniques. Elle prend la forme d'un plan de classement compris de l'ensemble des participants. → *Quelques bonnes pratiques de classement sont mentionnées en fin de ce document.*
- ✓ **Intégrer la correspondance avec l'extérieur** : prendre les mesures pour que la correspondance papier et électronique se retrouve, le plus automatiquement possible, dans le dossier du projet. Voir avec le service informatique pour une solution concernant le renvoi, classement et nommage automatique des courriels dans le dossier du projet.
- ✓ **Assurer un lien clair entre les parties dématérialisées et les parties matérialisées** : s'assurer que les parties matérialisées (papier etc.) sont bien classées et nommées de la même manière que le dossier électronique. Dans l'idéal, ces parties sont rassemblées dans un même lieu de stockage, gérées et contrôlées par les contributeurs chargés du contrôle des informations du projet.
- ✓ **Séparer la documentation scientifique extérieure de la production propre** : La documentation extérieure sert aux chercheurs pendant leur projet, mais n'appartient pas au projet proprement dit. Une liste bibliographique suffit pour les archives. C'est pourquoi, cette documentation doit être bien identifiée comme telle et être classée à part, bien séparée des documents produits pour qu'elle puisse être facilement éliminée ou recyclée à l'issue du projet.
- ✓ **Appliquer un choix restreint et pas trop « exotique » des formats des fichiers et données** : un format pérenne pour la conservation d'un document électronique permet d'accéder durablement aux informations contenues dans les fichiers, indépendamment d'un environnement système ou applicatif spécifique. Aussi, doit-on privilégier les formats dits « ouverts » car disposant d'une normalisation internationale. (Ex : pour les documents bureautiques, les formats recommandés sont : « .txt », « .jpg » ou « .pdf ».)
Il est fortement recommandé, avant le début du projet, de faire une évaluation technique et pratique pour la pérennité de lisibilité des formats de fichiers prévus, au moins pour les informations à conserver longtemps (en consultant le service informatique et le service des archives).
La question des formats lisibles et compatibles se pose aussi dans la perspective que toutes les informations du projet devront être exploitables par tous les participants du projet. Des formats trop « exotiques » peuvent ne pas être exploitables sur tous les environnements techniques (variables selon les organismes participants au projet).

Selon les pratiques de l'établissement, la **migration**¹ des fichiers vers des formats d'archivage peut être demandée auquel cas il faut s'assurer d'avance que toutes les compétences et outils de conversion sont disponibles.

Pour d'éventuels besoins de preuve, il faut s'assurer de la traçabilité des interventions sur les données (modifications, suppression, conversion, versions différentes etc.). Si la preuve doit être conservée à long terme, les enregistrements de traçabilité doivent être faciles à exploiter et récupérables en dehors du système de production des données.

Phase de finalisation

- ✓ **Vérification par le chef du projet de l'intégralité des données et informations dans le dossier/endroit unique d'enregistrement puis clôture officielle**

- ✓ **Définition des modalités de transmission pour archivage, des critères de tri et des durées de conservation** : Le moment de l'archivage n'a pas lieu avant la clôture du projet (l'archivage ne prend en compte que des données figées, donc qui n'ont plus vocation à être modifiées après archivage).

En cas de programmes pérennes, définir une politique d'archivage des données en continu, c'est-à-dire archiver à intervalles réguliers les données non susceptibles d'être modifiées ou bien les données telles qu'elles sont à un moment donné (« miroir »).

Ex : archiver les données de + 5 ans ; archiver les données disponibles dans une base de données tous les 31 décembre.

Pour le tri et les modalités de transmission pour archivage (support d'archivage, conversions, répertoire ou bordereau de versement à remplir etc.) il convient de consulter le service d'archives de l'organisme pour apprendre les procédures applicables.

→ *Voir quelques explications sur le tri ci-dessous.*

A défaut d'un dispositif d'archivage électronique, on peut songer à stocker les données à archiver dans un espace sécurisé par des droits d'accès restrictifs et séparé de l'espace de production ou bien de transmettre un CD/DVD en double exemplaire au service d'archives ou informatique.

Même dans ce cas, il est indispensable d'en avertir le service informatique et le service d'archives pour qu'ils apportent leur expertise sur les modalités du stockage pour assurer la sauvegarde régulière et la veille sur la pérennité des fichiers.

¹ = changement de format

Quelques conseils pour le nommage de fichiers informatiques

- **Longueur du nom d'un fichier : 20 caractères recommandés (espaces compris).**
- **Ordre des éléments du nom du fichier :**
 1. *Sujet principal du document* : Ex : *conseil administration, étude x, ...*
 2. *Typologie documentaire* : compte-rendu, rapport, note...
 3. *Date* : sous le format année, mois, jour. Les chiffres sont séparés par le tiret (« - ») ou le tiret bas (« _ »).
 4. *Version* : indiquée par l'abréviation « v » suivie par le numéro de la version. Version finale : « vf ».

→ Ex : CA_CR_2010-07-06_v2.doc
- **Pour séparer les éléments du nom d'un fichier, privilégier le tiret bas (« _ »)**
- **A éviter :**
 - Les phrases en guise de titre (verbes, adverbes, adjectifs, articles, prépositions, conjonctions, etc.)
 - Les majuscules, les accents et la ponctuation
 - Les caractères spéciaux ou les caractères réservés de l'environnement Windows (% , \$, § , * ; / , etc.)
 - Le nom du créateur du fichier inséré dans le titre
- **L'emploi des abréviations est limité à celles admises par l'établissement.**

Quelques bonnes pratiques de classement

- **Créer des dossiers et sous-dossiers liés aux différentes fonctions de l'étude en allant du général au particulier.**

Ex : Dossier « Etude xx » avec comme sous-dossiers :

 - Contexte de l'étude : protocole / questionnaire
 - Bases de données : base de données / documentation (dictionnaire de données, modèle de données)
 - Publication : rapport / note principal(e) / autres rapports / notes / articles
 - Dossier juridique : CNIL / convention
 - Correspondance / mail : *classement chronologique*
 - Réunions : *par type de réunions puis par classement chronologique*
- **Ne pas créer de dossiers portant le nom de la personne à l'origine du dossier ;**
- **Dans chaque dossier, distinguer dans des sous-dossiers distincts les versions préparatoires de la version finale. Les versions préparatoires peuvent rapidement être éliminées.**

Quelques explications sur les modalités de tri

Si elles n'existent pas déjà dans des documents réglementaires de l'organisme (**tableau de gestion des archives**) ou dans des **instructions de tri officielles**, les **règles de tri doivent être fixées en accord avec le service d'archives** de l'organisme et/ou avec le service public d'archives du périmètre géographique ou statutaire auquel appartient l'organisme (Archives municipales, départementales, nationales, etc.).

Pour définir les critères de tri, selon la finalité du projet, des opportunités et **obligations différentes de conservation** des informations s'imposent. Il convient de vérifier les obligations d'ordre légal, réglementaire, stratégique (politique de communication, d'archivage, de sécurité, système Qualité etc.), contractuel (partenaires, clients etc.), les besoins d'ordre scientifique à moyen et long terme ainsi que l'éventuel intérêt commercial.

Exemple : il peut n'y avoir aucune obligation légale de conserver les données brutes, par intérêt scientifique à long terme leur conservation s'impose tout de même.

En fonction des opportunités et obligations détectées, les informations différentes doivent être conservées selon des **règles parfois variées**.

Exemple : dans un même dossier de projet de recherche, le dossier financier doit être conservé pendant x années selon une obligation réglementaire, les rapports finaux sont à conserver sans limite de durée pour leur intérêt scientifique et historique.

Il est essentiel de fixer pour chaque information sa **durée de conservation**, le **moment de démarrage de cette durée** et le **sort final** (qui peut être soit la destruction, soit la conservation sans limite dans un service public d'archives) de l'information à l'issue de la durée fixée.

Il convient éventuellement de définir **sous quelle forme et pour quels publics** les informations doivent rester **disponibles** et accessibles pendant toute la durée de conservation.

Pour les organismes publics ou en charge d'une mission de service public : avant toute destruction (déchiquetage pour le papier, effacement des données électronique), un **visa de l'administration des archives** (Direction générale des patrimoines du Ministère de la Culture) est à obtenir moyennant un **bordereau d'élimination** à faire viser.