

Présentation de DoRANum



www.doranum.fr



Les objectifs



- Offre de ressources de **formation et d'auto-formation**
 - Déclinaison des modalités et des contenus selon :
 - ✓ Les métiers
 - ✓ Les disciplines
 - ✓ Les préférences, aptitudes et disponibilités individuelles
 - Ressources produites dans le cadre du programme et ressources externes
 - ✓ Importance des partenariats
- Sur la thématique des **données de la recherche**

Les public(s) cible(s)



- Enseignants-chercheurs
- Chercheurs
- Professionnels de l'IST et de soutien
 - Documentalistes
 - Bibliothécaires
 - Archivistes
 - Ingénieurs
- Doctorants
- Informaticiens

La « gouvernance »



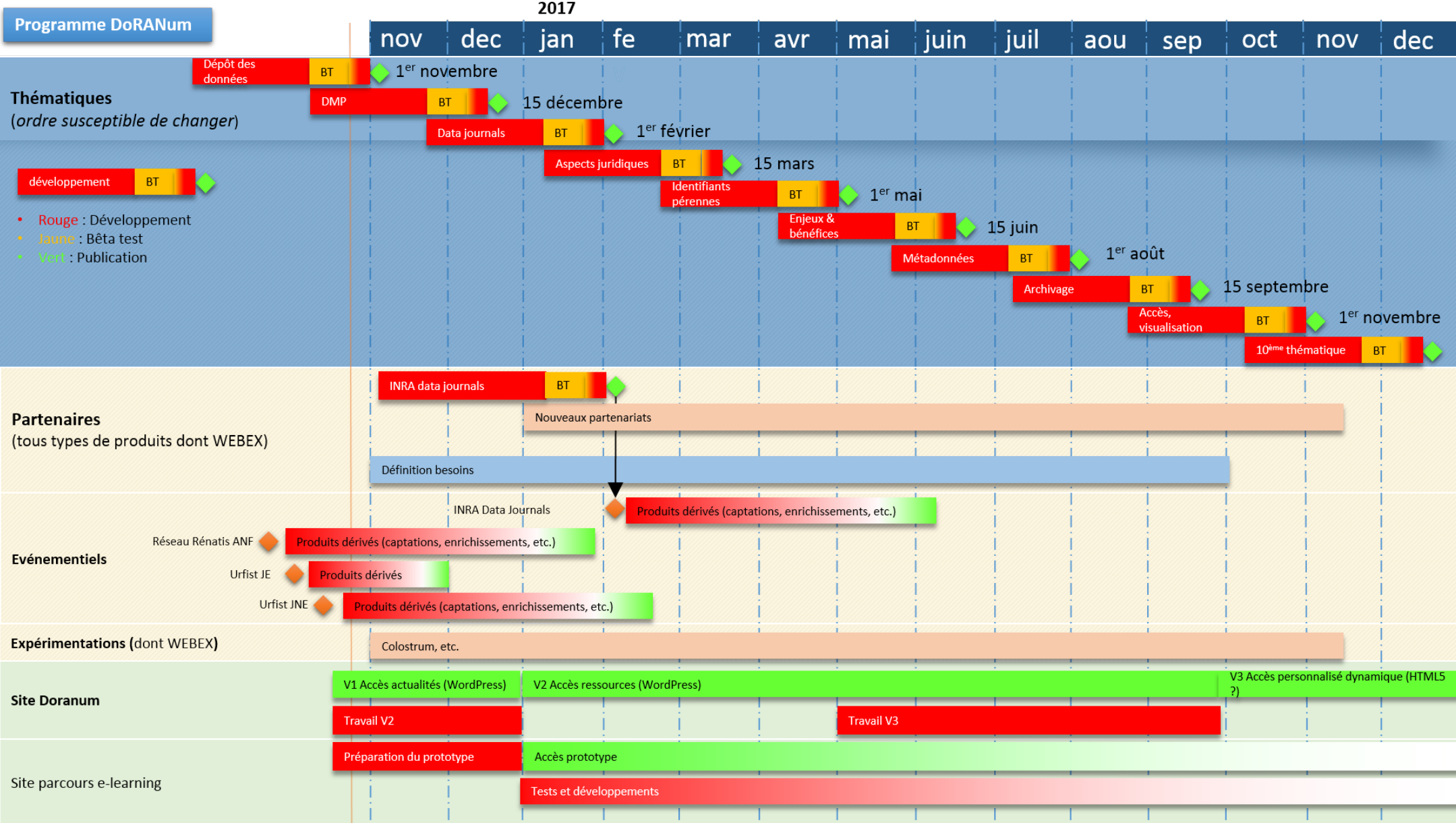
- Coopération Inist-CNRS et réseau des Urfist sous l'égide de BSN9
 - Deux ans (2016 – 2017)
 - Deux ingénieurs dédiés (Urfist de Nice et Inist à Nancy)
- Comité de pilotage: Inist-CNRS, réseau des Urfist, Dist-CNRS, BSN10 & BSN9
- Groupe de travail d'une vingtaine de personnes (CP + ingénieurs + ministère + partenaires + experts...)

Les partenariats



- 3 propositions de partenariats (non limitatives)
 - *partenaires désireux de produire des modules de formation à distance*
 - *partenaires qui réalisent ou prévoient de réaliser des formations*
 - *partenaires-relais et bêta-testeurs qui seraient volontaires pour tester et diffuser les modules de formation à distance produits par DoRANum*
- Aujourd'hui
 - *des organismes de recherche (Inra, CEA...)*
 - *des Universités (Nice Sophia A. ...)*
 - *Un consortium (Couperin)*
 - *Des MSH ("Ange Guépin" Nantes...)*
 - *...*

Le macroplanning



La JNE 2016 du réseau des Urfist, « Partage des données de la recherche », en « Sketchnotes » par Magalie Le Gall

Magalie Le Gall a réalisé des **sketchnotes** (notes graphiques, mi cartes heuristiques, mi infographies, réalisées à la volée) épatantes lors de la **journée nationale d'études du réseau des Urfist**: « Partage des données de la recherche : quels impacts ? » qui a eu lieu à Paris le 29 septembre dernier. Elle les a postées **sur Facebook** (accès réservé) avec la mention suivante: « Ce n'est pas impossible qu'il y ait des inexactitudes ou des contresens, n'hésitez pas à me les signaler si vous en voyez ! 😊 ».

Suite à notre demande, elle en permet la réutilisation sous licence **CC-By**.



Le programme détaillé se trouve sur le site de la JNE. Les captations des interventions sont en cours de traitement.

ANF « Participer à l'organisation du management des données de la recherche, gestion de contenu et documentation des données » – 3 au 6 juillet 2017

Les actualités



Les thématiques



Enjeux et bénéfices
Plan de gestion de données
ID pérennes et DOI
Métadonnées
Aspects juridiques et éthiques

Stockage et archivage
Dépôt des données et
entrepôts
Accès et visualisation
Data journals et data papers

ZOOM SUR...

100 jours de durée de vie pour une URL en moyenne

selon Brewster Kahle, le fondateur d'[Internet Archive](#). Pour assurer un accès pérenne à ses ressources sur le web, une des solutions est l'attribution d'un DOI !

EN SAVOIR PLUS

THEMATIQUES



ENJEUX - BENEFICES

Pourquoi partager les données ? Qu'est-ce que l'Open Science ?



DMP - PLAN DE GESTION DES DONNEES

Pourquoi et comment rédiger un plan de gestion des données ?



ID. PERENNES - DOI

Comment citer un jeu de données ?



METADONNEES - FORMATS

Comment décrire mes données ?



ASPECTS JURIDIQUES ET ETHIQUES

Que puis-je partager ? Que puis-je réutiliser ?



DEPOT - ENTREPOTS

Où et comment déposer mes données ?



ACCES - VISUALISATION

Où et comment extraire les données qui m'intéressent ?



DATA PAPERS - DATA JOURNALS

Pourquoi et comment publier mes données comme un article scientifique ?



STOCKAGE - ARCHIVAGE

Quelles données conserver à long terme et comment ?



En bref



La fiche synthétique

Fiche 2 min PDF



La minute « Entrepôt »

Video 1 min PDF

Le dépôt des données de recherche répond à une logique de partage et nécessite une gestion rigoureuse des données pour que celles-ci soient accessibles et réutilisables dans le temps.

Consultez les sujets développés pour cette thématique et n'hésitez pas à laisser vos commentaires sur les pages des produits !

L'essentiel



Du choix de l'entrepôt au dépôt des données

Infographie 10 min PDF



Les critères de choix d'entrepôt

Fiche 5 min PDF



Vérifier ses données

Checklist 10 min PDF



Tutoriel Comment déposer ses données dans Zenodo

Tutoriel 2 min PDF

Pour aller plus loin

e-learning

Suivez cette petite formation regroupant toutes les ressources sur la thématique du dépôt des données de recherche.

C'EST PARTI !

Autres ressources



Le glossaire



La FAQ

Les niveaux de formation

En bref

L'essentiel


Pour aller plus loin

En bref

En bref


Le dépôt des données de recherche répond à une logique de partage et nécessite une gestion rigoureuse des données pour que celles-ci soient accessibles et réutilisables dans le temps.

Consultez les sujets développés pour cette thématique et n'hésitez pas à laisser vos commentaires sur les pages des produits !



La fiche synthétique

Fiche 2 min PDF



La minute « Entrepôt »

Vidéo 1 min PDF

Testez vos connaissances !

QUESTION 1/4 25%



Un entrepôt de données permet...

- d'effectuer des sauvegardes régulières des jeux de données
- de stocker des jeux de données, d'y accéder et de les réutiliser
- d'accéder à un espace collaboratif de travail sur ses jeux de données

Montrer la réponse

Fiche
synthétique
Minute (vidéo
très courte)
Quiz

L'essentiel

L'essentiel



Le data mining



Présentation



3 min



Les bases de données
et les logiciels



Infographie



2 min

Tutoriels multimédias
Présentations
Infographies
Fiches synthétiques de
2^{ème} niveau
Guides
Checklists
...

Pour aller plus loin

Pour aller plus loin

e-learning

Suivez cette petite formation regroupant toutes les ressources sur la thématique du dépôt des données de recherche.

C'EST PARTI !

Pour aller plus loin

Hélène SKRZYNIAK



L'open data : quelle protection ?
Quelles obligations pour le chercheur ?

L'open data : quelle protection ?
Quelles obligations pour le chercheur
?

Exposé **Hélène SKRZYNIAK**

41 min

Cours
Vidéos
(formations...)

...

Autres ressources

Autres ressources



Le glossaire

VEZ ENRICHIR ICI
VOTRE VOCABULAIRE SUR
LES DONNÉES DE LA
RECHERCHE.



La FAQ

UNE QUESTION ? POSEZ
LA À NOS PARTENAIRES
OU CONSULTEZ LES
RÉPONSES DÉJÀ POSTÉES.

Exemples de ressources pédagogiques en détail



Fiches synthétiques

Vidéos

Tutoriels scénarisés multimédias

Infographies

Présentations

Quiz

...

En bref

Le partage des données va de pair avec une bonne gestion des données. En effet, pour pouvoir être diffusées, les données doivent avoir été au préalable décrites, documentées, conservées dans un format réutilisable, etc. Mais pourquoi partager ses données ? Quels sont les enjeux pour les acteurs de la recherche, la science en général et la société ?

Consultez les sujets développés pour cette thématique et n'hésitez pas à laisser vos commentaires sur les pages des produits !



La fiche synthétique

Fiche 2 min PDF



La minute
« Validation et réutilisation
des données »

Video 1 min PDF

Testez vos connaissances !

QUESTION 1/4

25%



Les projets de recherche financés par de l'argent public doivent obligatoirement partager ouvertement toutes leurs données.

- Tout à fait
- C'est faux

Montrer la réponse

L'essentiel



Les acteurs impliqués dans la gestion des données de la recherche

Présentation 5 min PDF



Les freins et défis liés à la gestion et au partage des données

Sondage 2 min PDF

ENJEUX ET BÉNÉFICES

POUR LES CHERCHEURS

De nouvelles exigences et opportunités

- Le partage des données peut être **obligatoire** pour obtenir une subvention ou pour publier un article.
- Il permet aussi au chercheur de **valoriser ses données** et d'**accroître sa visibilité**.

Exemples

Les bénéficiaires d'un financement H2020 doivent planifier le dépôt de leurs données...

Les données sont citables et la paternité du chercheur est reconnue.

POUR LA SCIENCE

De nouvelles perspectives

Le partage des données permet aux chercheurs de s'appuyer sur un socle **plus important** et ouvre la voie à de **nouvelles méthodes d'investigations**.

Utiliser des bases de données de référence, approcher de différentes façons un même jeu de données...

Une meilleure validation

- L'**accès aux données améliore les conditions de validation des articles**. Aussi le dépôt des données est de **plus en plus demandé par les comités de lecture**.
- La réutilisation des données par d'autres chercheurs peut amener au **réexamen des résultats initiaux**.

POUR LA SOCIÉTÉ

La réutilisation des données permet un **milleur usage de l'argent public** et dynamise les relations entre la recherche et la société.

Implication des citoyens dans le cadre des [Sciences participatives](#)

Le saviez-vous ?

La croissance très importante du nombre de données dans le domaine scientifique mais aussi en général ([Big data](#)) présente à la communauté scientifique des **champs et des exigences nouvelles** ([Data Driven Science](#), [Humanités Numériques](#))



POUR LES CHERCHEURS

- Le partage des données peut être une condition pour l'obtention du financement de projets scientifiques. Par exemple, pour obtenir un financement dans le cadre d'Horizon 2020, il est nécessaire de préparer un [Plan de gestion des données](#)
- Pour la publication d'un article, certains éditeurs recommandent, voire exigent, l'accès aux données. Par exemple, l'éditeur [PLOS ONE](#) oblige les auteurs d'un article à rendre accessibles toutes les données sous-jacentes aux résultats rapportés dans l'article soumis.
- Les chercheurs peuvent mieux promouvoir leurs recherches et être cités, car les données rentrent dans le processus d'édition scientifique ([dépôt de données](#), [publication de data papers](#)).
- Les auteurs des jeux de données sont [identifiables](#), ce qui facilite les potentielles collaborations entre chercheurs.

POUR LA SCIENCE

- Rendre ses données disponibles offre une meilleure garantie contre la fraude scientifique.
- Partager ses données nécessite d'adopter de bonnes pratiques de gestion des données (décrire les données, les documenter, les pérenniser...) ce qui améliore la qualité du travail de recherche.
- Le coût engendré par la création, la collecte, le traitement des données peut être très élevé. Réutiliser des données déjà existantes plutôt que de les recréer permet de rentabiliser la recherche, d'accélérer l'innovation et le retour sur investissement dans la Recherche et Développement.
- La constitution de bases de données permet la fouille de données ([Text Data Mining](#)), de les extraire, de les recouper et d'en construire des visualisations. Ces nouveaux procédés facilitent l'initiative de nouvelles recherches et leur interdisciplinarité.
- Le déluge des données numériques ([Big Data](#)) impacte la manière de faire de la recherche scientifique. On parle de [Data Driven Science](#), une démarche qui automatise les découvertes en exploitant la puissance des ordinateurs pour trouver des corrélations parmi de grandes quantités de données.

POUR LA SOCIÉTÉ

- Les recherches financées par l'argent public doivent pouvoir bénéficier à tous. Ouvrir les données rend la recherche plus transparente, renforce la confiance des citoyens et leur permet de s'impliquer (dans le cadre des [sciences participatives](#) par exemple).
- Les données générées par l'[Open Data](#) et le [Big Data](#) offrent un terrain d'exploitation à la recherche scientifique, qui en retour peut éclairer la société sur ses évolutions les plus récentes.



LA MINUTE

"VALIDATION ET RÉUTILISATION DES DONNÉES"

Les structures institutionnelles

Les communautés scientifiques

Les éditeurs

**Les acteurs
impliqués
dans la GDR**

Les aspects juridiques et éthiques

En bref

Ce sujet un peu rébarbatif n'en est pas moins important. Connaissez-vous les critères de communication et de diffusion de vos données ? Si vous réussissez à traiter les aspects juridiques, comment traiter les aspects éthiques ? Changez-vous la diffusion de vos données pour ne pas porter atteinte aux sujets étudiés ?

Consultez les sujets développés pour cette thématique et n'hésitez pas à laisser vos commentaires sur les pages des produits !



Testez vos connaissances !

QUESTION 1/4 25%

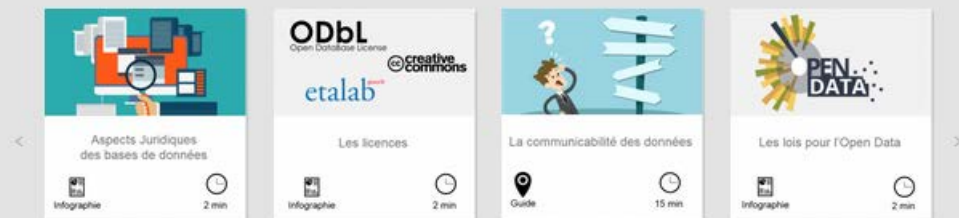


Est-ce que les licences sont obligatoires pour diffuser dans l'Open Data ?

Oui
 Non

Montrer la réponse

L'essentiel



Pour aller plus loin



Aspects Juridiques & Ethiques

Le mouvement d'ouverture de la science (Open Science ou Open Research) doit s'accompagner d'une réflexion qui précisera les conditions de partage dans le respect de la propriété intellectuelle. La nouvelle loi pour une République numérique, dite Loi Lemaire, offre un nouveau cadre juridique pour tenter de répondre davantage à l'ouverture des données de la recherche. Ce nouveau cadre précise les conditions de communication et de diffusion.

Le partage des données

Le cadre juridique de l'Open Data

L'Open Data a trouvé ses bases dans la loi initiale dite CADA (Commission d'Accès aux Documents Administratifs). La loi Valter est relative à la gratuité et aux modalités de réutilisation des informations du secteur public. La loi dite Lemaire apporte une meilleure protection pour le partage et la réutilisation des productions scientifiques.

Notes

Les lois dites CADA, Valter et Lemaire sont maintenant codifiées dans le code des relations entre le public et l'administration.

La dimension éthique

Le respect de la vie privée, la propriété intellectuelle, la qualité et l'intégrité des données sont des dimensions éthiques de la gestion des données. Etre informé de l'évolution des règles juridiques est aussi très important pour un partage responsable des données utilisées.

L'adaptation de l'Open Data pose la question de compétences nouvelles qui répondront au besoin d'information des chercheurs. Le COMETS¹ préconise la création de comités d'éthique sur les données de recherche par discipline ou établissement.

Le saviez-vous ?

Les conditions de diffusion dans l'Open Data sont à vérifier par l'analyse de la nature juridique et éthique des données. L'outil d'analyse présenté au verso permet de déterminer si les données sont communicables et diffusables. Les données doivent aussi être considérées comme achevées² et être réalisées dans le cadre de la mission de service public des établissements.

Les droits d'auteur et la diffusion

Code de la propriété intellectuelle

Depuis 2006 (loi DADVSI), les agents publics sont titulaires des droits patrimoniaux et moraux de leur œuvre. Cependant une distinction importante sépare les différentes catégories d'agent.

Les chercheurs et enseignants-chercheurs, catégories d'agents non soumis à autorité hiérarchique, conservent la plénitude de leurs droits d'auteur, alors que les autres catégories ne conservent que leur droit de paternité, cédant à l'employeur public les droits patrimoniaux et les autres droits moraux.

Notes

Les droits moraux sont constitués des : droit de paternité, droit de 1^{ère} divulgation, droit au respect de l'œuvre, droit de retrait ou de repentir.

Les droits patrimoniaux sont constitués des : droit de représentation, droit de reproduction,

Les licences à mettre en œuvre

L'analyse des jeux de données détermine la possibilité de diffuser dans le cadre de l'Open Data.

La licence choisie par l'auteur engage le ré-utilisateur à respecter l'intégrité des données, à faire mention de la source des données et à porter l'indication de la date de dernière mise à jour.

Les licences les plus courantes sont de type Creative Commons, ODbL (Open Database Licence), Etalab (initiative gouvernementale). Le guide d'analyse mentionné au verso fournit une aide sous forme de tableau comparatif de critères.

Un décret de la loi dite Lemaire va venir préciser la liste des licences utilisables.

¹ COMETS : Comité d'éthique du CNRS

² Les documents préparatoires contribuant au processus de recherche ne sont pas considérés comme données achevées.

Un outil pour l'analyse des données

La communicabilité des jeux de données est déterminée par l'analyse de la nature juridique et éthique de l'ensemble des données. Le guide « Ouverture des données de recherche, analyse du cadre juridique en France » propose en page 29, un logigramme de communicabilité des données.

- Le logigramme est divisé en trois branches principales :



- La progression de lecture peut être effectuée selon :

Types juridiques :

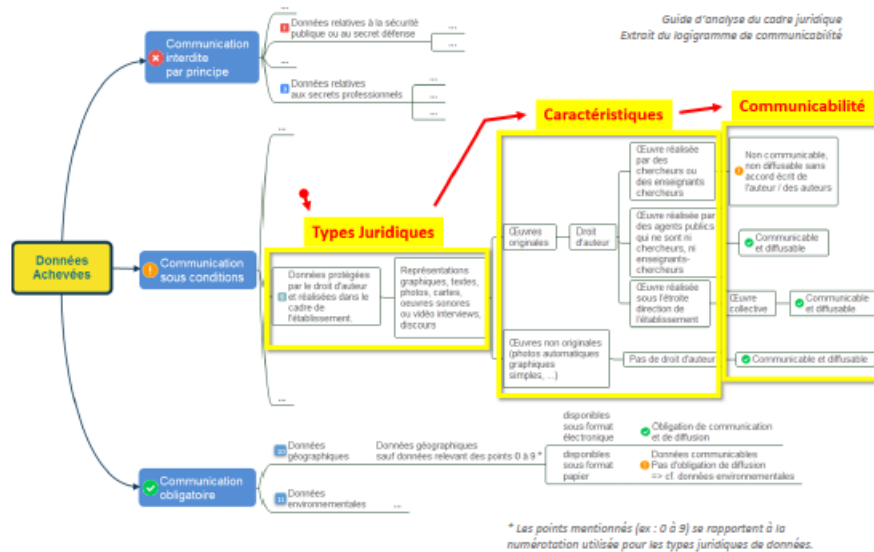
La donnée est-elle une photo, est-elle relative aux secrets professionnels, est-elle géographique, etc. ? Localiser la branche qui se rapproche du type de la donnée concernée.

Caractéristiques :

La donnée provient-elle d'une unité protégée, est-ce une marque, est-elle médicale, etc. ? Poursuivre dans les embranchements pour préciser les caractéristiques de la donnée concernée.

Communicabilité :

La terminaison des embranchements précise si la donnée est communicable, diffusable, non communicable, non diffusable, avec l'accord de..., sous conditions d'un contrat, avec l'autorisation de...

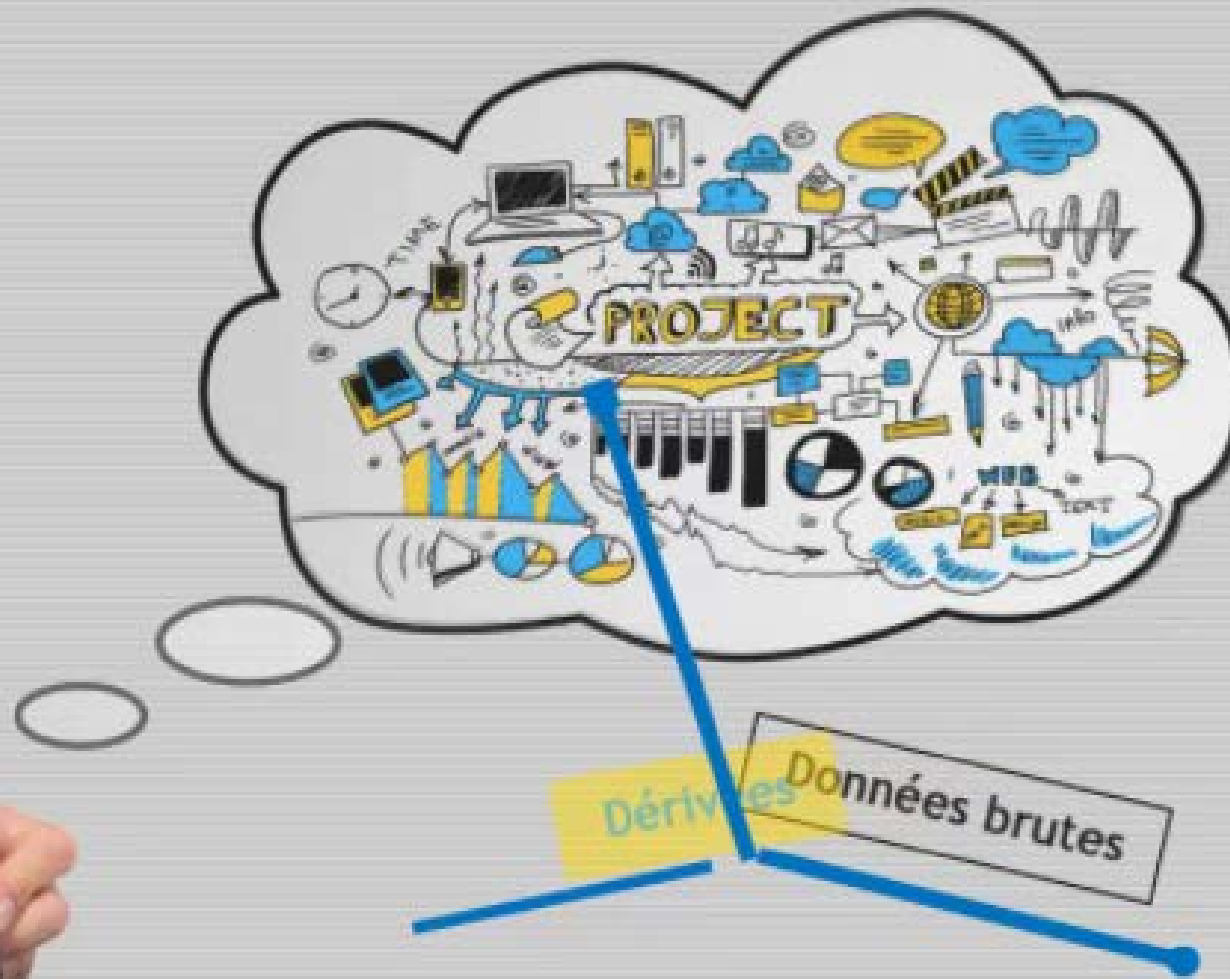


* Les points mentionnés (ex : 0 à 9) se rapportent à la numérotation utilisée pour les types juridiques de données.



N'hésitez pas à prendre conseil auprès du juriste associé à votre établissement.

Les Aspects Juridiques et Ethiques



Les droits d'auteur

Depuis 2006, la loi dite DADVSI, loi relative au **Droit d'Auteur** et aux **Droits Voisins** dans la **Société de l'Information** a intégré les agents publics dans le code de la propriété intellectuelle.

Une distinction existe entre les catégories d'agents soumis ou non au contrôle préalable de l'autorité hiérarchique.

Les chercheurs et les enseignants-chercheurs, non soumis à autorité hiérarchique, bénéficient de la plénitude de leurs droits d'auteur.

Les agents soumis à autorité hiérarchique ne conservent que les droits de paternité de leur œuvre, les droits patrimoniaux et les autres droits sont exercés par l'employeur public.



[Visualiser en plein écran](#)



[Visualiser en plein écran](#)

Communicable et diffusable dans l'Open Data

Œuvre réalisée par des chercheurs ou des enseignants-chercheurs

non communicable, non diffusable sans accord écrit de l'auteur / des auteurs

Œuvre réalisée par des agents publics qui ne sont ni chercheurs, ni enseignants-chercheurs

communicable et diffusable

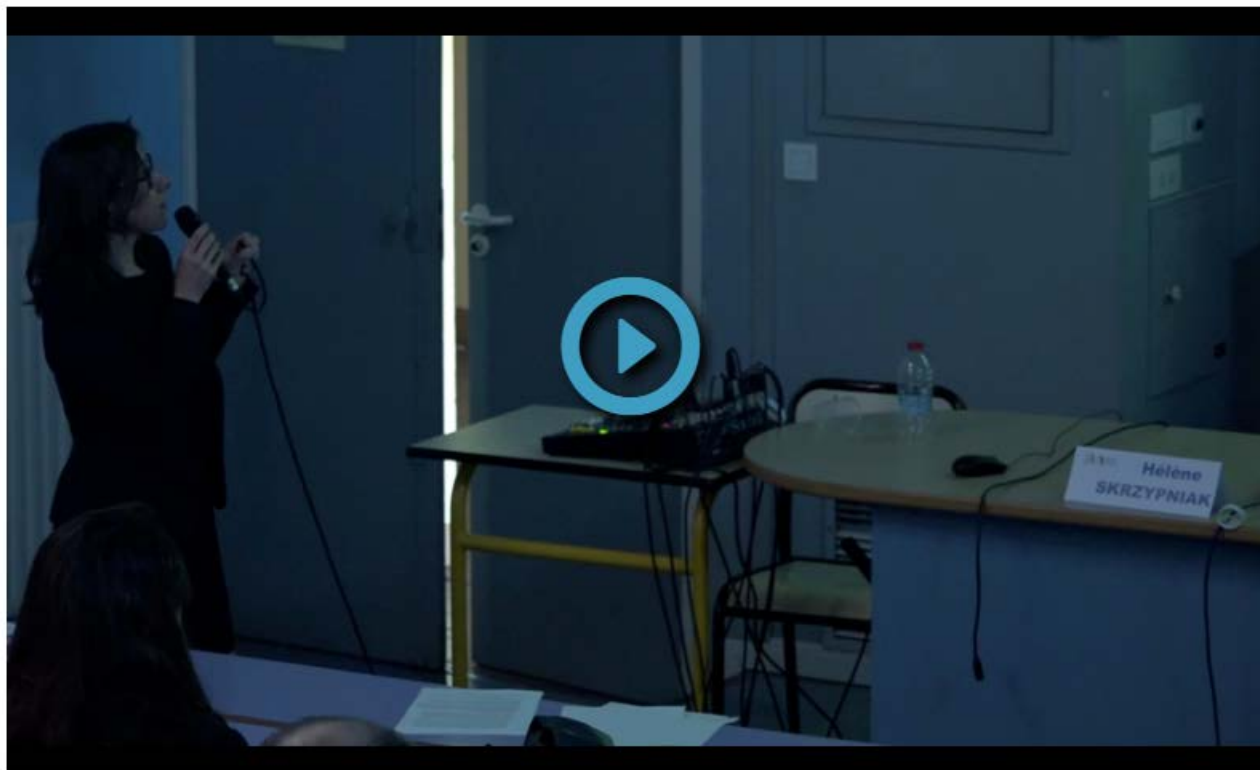
Œuvre réalisée sous l'étroite direction de l'établissement => oeuvre collective

communicable et diffusable

Exposé

L'open data : quelle protection ? Quelles obligations pour le chercheur ?

Exposé Hélène SKRZYNIAK, enseignante-chercheuse en droit de la propriété intellectuelle (Irdap, université de Bordeaux)



L'open data : quelle protection? Quelles obligations pour le chercheur

Hélène SKRZYNIAK

13/04/2017

4,0 x 190,5 mm

Le plan de gestion de données : DMP

En bref

Le plan de gestion des données est un outil de gestion. Il se présente sous forme d'un document structuré en rubriques. Il a pour objectif de synthétiser la description et l'évolution des jeux de données de votre projet de recherche. Il prépare le partage, la réutilisation et la pérennisation des données.

Consultez les sujets développés pour cette thématique et n'hésitez pas à laisser vos commentaires sur les pages des produits !



La fiche synthétique

Fiche 3 min PDF



La minute « DMP »

Vidéo 2 min

Testez vos connaissances !

QUESTION 1/4

25%




Le DMP (Plan de gestion de données) est un document qui décrit la progression d'un travail de recherche.

- Vrai
 Faux

Montrer la réponse

L'essentiel




La description des données

Vidéo 3 min



L'origine des données

Vidéo 1 min



Outil de rédaction DMP

Vidéo 1 min



Data Management Plan (DMP)

Le Data Management Plan ou Plan de gestion de données est un document synthétique qui aide à organiser et anticiper toutes les étapes du cycle de vie de la donnée. Il explique pour chaque jeu de données comment seront gérées les données d'un projet, depuis leur création ou collecte jusqu'à leur partage et leur archivage.

Une bonne pratique de recherche

De plus en plus utilisée

Le contexte politique va en faveur de l'Open Data. Recommandé par les financeurs, le DMP devient l'outil de gestion incontournable des projets de recherche.

Pour optimiser la gestion de vos données

Exemples

L'Europe, les Etats-Unis ou encore l'Australie ont adopté le DMP. Les projets rentrant dans le pilote du programme Horizon 2020 doivent obligatoirement en établir un.

Orientée libre accès

Le DMP est très lié au principe du libre accès aux données de recherche. En fonction de votre choix et de vos contraintes en matière de partage, des critères sont à définir.

Pour vous poser les bonnes questions

Quelles métadonnées renseigner ? Faut-il des outils spécifiques pour accéder aux données ? Quelle licence correspond à vos souhaits et/ou contraintes de partage ? Etc.

Le saviez-vous ?

Un DMP peut être établi aussi bien dans une optique de partage des données que pour des données en accès restreint ou fermé, total ou partiel. Le DMP mentionnera dans ce cas les raisons de non partage.

La réalisation d'un DMP

Suivant un calendrier

Le DMP est un document évolutif. Des mises à jour et des livrables précis peuvent être définis selon les financeurs et/ou projets.

Pour respecter les échéances

Exemples

Mise à jour lors de la production d'un nouveau jeu de données, lors du dépôt d'une demande de brevet, etc.

Au moyen d'outils dédiés

Des outils existent pour vous aider à produire votre DMP. Ils sont gratuits après création d'un compte.

Pour vous guider

DMP OPIDoR, instance française de DMPonline du Digital Curation Centre (DCC)

Modèle de DMP

Il n'existe pas de trame unique, mais de nombreux modèles de DMP ont toutefois été établis par des organismes, instituts, financeurs à destination de leurs utilisateurs : par exemple le modèle de [DMP H2020 établi par la Commission européenne](#) ou celui du [Cirad](#).

Les outils d'aide à la rédaction comme [DMP OPIDoR](#) donnent accès à des modèles et facilitent la rédaction du DMP grâce à des guides et des exemples personnalisés..

Le degré de précision et l'agencement des champs à compléter peuvent varier d'un modèle à l'autre. Dans tous les cas, il s'agit d'une série de questions organisées en grands chapitres. On retrouve systématiquement les mêmes éléments principaux qui sont :



La minute DMP



La description des données

Le DMP intègre nécessairement la description de vos données, sans méthodologie prédéfinie. Trois pistes de classifications descriptives vous sont ici proposées : la description des données en fonction de leur source, de leur forme et de leur stabilité. Différentes approches sont cumulables pour offrir une compréhension complète des données.



Le dépôt des données

En bref

Le dépôt des données de recherche répond à une logique de partage et nécessite une gestion rigoureuse des données pour que celles-ci soient accessibles et réutilisables dans le temps.

Consultez les sujets développés pour cette thématique et n'hésitez pas à laisser vos commentaires sur les pages des produits !



Testez vos connaissances !

QUESTION 5/4

25%



Un entrepôt de données permet...

- d'effectuer des sauvegardes régulières des jeux de données
- de stocker des jeux de données, d'y accéder et de les réutiliser
- d'accéder à un espace collaboratif de travail sur ses jeux de données

Montrer la réponse

L'essentiel



Vérifier ses données



Déposer ses données dans Zenodo



Le dépôt des données en 5 questions



Du choix de l'entrepôt au dépôt des données



Pour aller plus loin

e-learning

Suivez cette petite formation regroupant toutes les ressources sur la thématique du dépôt des données de recherche.

C'EST PARTI !



DÉPOSER SES DONNÉES

✓ OÙ DÉPOSER SES DONNÉES ?

Dans un **entrepôt**. Mais il existe des milliers d'entrepôts ! Disciplinaires, multidisciplinaires, propres à un éditeur : alors comment trouver l'entrepôt qui correspond à vos besoins ?

1 Suivez les recommandations

Un entrepôt vous est recommandé par l'une des parties prenantes de votre projet (votre financeur, éditeur, institution, etc.) ?

➤ Renseignez-vous !

Exemples

En génomique, la communauté scientifique utilise notamment [Uniprot](#) ou [GenBank](#) comme base de référence.

2 Filtrez par critères

Aucun entrepôt ne vous est recommandé ? Il vous faut donc en trouver un adapté à vos besoins.

➤ Posez-vous les bonnes questions !

Souhaitez-vous un entrepôt ciblé sur votre discipline ? Est-ce que l'entrepôt accepte les formats de vos fichiers ? Etc.

Le saviez-vous ?

Il existe des **annuaires** (ou répertoires) qui peuvent vous aider à filtrer votre recherche d'entrepôt : [re3data](#), [OAD](#), [OpenDOAR](#), etc.

✓ COMMENT DÉPOSER SES DONNÉES ?

3 Vérifiez vos données !

En effet, une préparation préalable de vos données est nécessaire pour pouvoir les déposer dans de bonnes conditions.

Vous avez choisi une licence de diffusion de vos données, anonymisé les données à caractère personnel, etc.

4 Remplissez le formulaire de dépôt

Quand tout est prêt, connectez-vous sur le site de l'entrepôt, remplissez le formulaire de dépôt et exportez vos fichiers.

Fiche synthétique – 2016
doranum.fr

L'essentiel



1) Suivez les recommandations

- Vos travaux de recherche rentrent dans le cadre d'un **projet national ou international** ?

Exemple : La politique de données du programme de recherche international [MERMeX](#) stipule que les données collectées seront recueillies dans [MISTRALS database](#) pour améliorer, faciliter et sécuriser leur disponibilité.

- Votre **institution** s'organise par rapport au dépôt des données ?

Exemple : L'[Institut Laue-Langevin](#) a mis en place une [politique de partage des données](#) et dispose d'un portail dans lequel les chercheurs de l'institut peuvent déposer leurs données.

- Il existe des pratiques spécifiques à votre **domaine scientifique** ?

Exemple : En génomique, les bases de données comme [Uniprot](#), [GenBank](#) sont aujourd'hui des bases de référence.

- Votre **éditeur** a une politique concernant le partage des données ?

Exemple : L'éditeur [PLOS ONE](#) oblige les auteurs d'un article à rendre accessibles toutes les données sous-jacentes aux résultats rapportés dans l'article soumis.

- Votre **financeur** vous donne des recommandations ?

Exemple : La Commission Européenne donne des [lignes directrices](#) aux chercheurs bénéficiant de financement dans le cadre d'Horizon 2020. Les bénéficiaires doivent déposer les données nécessaires à la validation des résultats présentés dans les publications et les autres données citées dans le plan de gestion des données. Le choix de l'entrepôt reste du ressort du déposant (l'annuaire [re3data](#) est recommandé et l'entrepôt [Zenodo](#) est également cité).

2) Filtrez par critères

Quels **types de données** sont acceptés ?

Pourrez-vous **modifier** les données déposées et accéder aux anciennes versions ?

Peut-on faire un **lien** entre les données et les publications associées ?

Une **licence** est-elle imposée ? Vous convient-elle ?

Souhaitez-vous que l'entrepôt s'engage à préserver l'accès aux données sur le **long terme** ?

Y a-t-il des **statistiques** d'utilisation, de consultation, de téléchargement ?

Y a-t-il un **coût** pour le dépôt ?

Quels **formats** sont acceptés ?

Pourrez-vous restreindre l'**accès** à vos données ?

Mettre un **embargo** ?

Quel **type d'entrepôt** privilégieriez-vous ? Disciplinaire ? Multidisciplinaire ? Certifié ?

Un **identifiant pérenne** est attribué aux données ?

Souhaitez-vous que l'**hébergement du serveur** de l'entrepôt soit localisé en France ?

Le site est-il **simple d'utilisation** ? Y a-t-il des **aides** proposées ?

Les réponses à ces questions peuvent être trouvées sur les **pages de politique des entrepôts** ou être **filtrées dans des annuaires** d'entrepôt.

3) Vérifiez vos données !

Les **principes éthiques** sont respectés

Les fichiers sont **organisés et nommés** de façon explicite

Les **modalités d'accès** sont définies

Une **licence** est attribuée aux données

Le **volume** des fichiers ne dépasse pas la taille maximale

Les **droits de diffusion** sont vérifiés

Les fichiers sont dans des **formats pérennes et ouverts**

Les jeux de données à partager sont **sélectionnés**

Les données sont **décrites et documentées**

Un **identifiant pérenne et unique** est attribué

Fiche synthétique – 2016
doranum.fr

Aide-mémoire



La minute entrepôt

Entrepôts

Annuaire



MIXTE = Articles & Données



Critères de qualité



Checklist avant de déposer ses données

- Les principes éthiques sont respectés
 - Vigilance à porter sur les données à caractère personnel
 - Se référer à la CNIL
- Les droits de diffusion des données sont vérifiés / obtenus
 - Vigilance à porter sur l'interdiction de diffusion de certaines données (secret professionnel...)
 - Obtenir la permission d'autres collaborateurs de diffuser les données
 - Se référer à l'accord de consortium
- Les modalités d'accès sont définies
 - Choisir un accès ouvert ou restreint
 - Choisir une période d'embargo ou non
- Une licence appropriée est attribuée aux données
 - Choisir une licence de diffusion
- Les jeux de données à partager sont sélectionnés
 - Structurer et agréger les données en jeux de données cohérents
- Les fichiers sont organisés et nommés de façon explicite
 - Créer des conventions de nommage des fichiers
- Les fichiers sont dans des formats pérennes et ouverts
 - Utiliser des formats de fichier acceptés par l'entrepôt
 - Privilégier les formats ouverts ou largement répandus
- Le volume des fichiers ne dépasse pas la limite autorisée
 - Se référer à la taille maximale autorisée par l'entrepôt
- Les données sont décrites et documentées
 - Utiliser un standard de description
 - Fournir une documentation, au minimum un fichier Readme
 - Fournir si besoin un dictionnaire de données
- Un identifiant pérenne et unique est attribué aux données
 - Se référer à son institution ou à l'entrepôt pour l'attribution

A retenir

Déposer ses données demande de les avoir préparées au préalable. La liste ci-dessus vous indique les points essentiels à vérifier pour pouvoir déposer vos données sans problème.

Fiche synthétique – 2016

doranum.fr



Tutoriel

Comment déposer ses données dans l'entrepôt Zenodo



Les data papers

En bref

Il existe plusieurs moyens pour publier ses données. Parmi eux, les data papers ont l'avantage d'être des articles à part entière, suivant la même logique de processus éditorial que les articles scientifiques classiques.

Consultez les sujets développés pour cette thématique et n'hésitez pas à laisser vos commentaires sur les pages des produits !



Testez vos connaissances !

QUESTION 1/4

25%



Un data paper est

- Un article de recherche
- Un tableau recensant toutes les données créées lors d'un projet de recherche
- Un article décrivant des jeux de données

Montrer la réponse

L'essentiel



PUBLIER UN DATA PAPER

Le data paper (data article, data descriptor) est une **publication qui décrit un jeu de données scientifiques**, notamment à l'aide d'informations structurées appelées métadonnées. Contrairement aux articles de recherches classiques, les data papers fournissent une voie formalisée au **partage des données** plutôt que tester des hypothèses ou présenter de nouvelles analyses.

POURQUOI PUBLIER UN DATA PAPER ?

- Pour **communiquer** sur l'existence des données et permettre de les **trouver** ;
- Pour **créditer** les auteurs (reconnaissance, référence citable) et **valoriser** les données ;
- Pour faciliter la **réutilisation** des données (en les rendant intelligibles).

OÙ PUBLIER UN DATA PAPER ?

- Dans un **data Journal**, revue dédiée à ce type de publication ;
- Dans une **revue classique** qui publie des data papers en plus des articles classiques.

QUE CONTIENT UN DATA PAPER ?

Les data papers ont une structure très variable selon les revues. Ils ont cependant des **composantes communes** :

Une partie descriptive

Exemples

- Éléments communs aux articles classiques ; *Titre, résumé, mots clefs, etc.*
- Éléments spécifiques liés aux données. *Métadonnées, réutilisation, etc.*

L'accès aux données

Elles peuvent être **Intégrées** dans l'article, publiées sous forme de **matériel supplémentaire** (supplementary data), ou être **déposées dans un entrepôt**. Dans ce dernier cas, l'identifiant des données (exemple : le DOI) permet d'établir le lien du data paper vers les données.

COMMENT RÉDIGER UN DATA PAPER ?

Appliquez les instructions aux auteurs et utilisez les articles déjà publiés dans la revue cible comme modèles. Notez aussi que :

- Certaines revues proposent leurs modèles ; *Le [template de Data in Brief](#).*
- D'autres proposent des outils de rédaction en ligne ; *L'outil [Pensoft Writing Tool](#).*
- D'autres autorisent la soumission à partir des plateformes externes. *[Scratchpads](#), [GBIF](#).*

Fiche synthétique – 2017
doranum.fr

L'essentiel



Fiche synthétique – 2017
doranum.fr

Aide-mémoire



Selon la revue et ses exigences, le contenu et la taille du data paper peuvent varier considérablement entre une **forme très synthétique** et un **article très complet**. Voici un exemple de structure.

EXEMPLE DE DATA PAPER DANS LE DATA JOURNAL "DATA IN BRIEF"



Data in Brief
Volume 10, February 2017, Pages 198–201



Open Access

Titre : Le titre doit se concentrer sur les données spécifiques partagées, ce n'est pas un titre d'article de recherche

Auteur(s) : Nom, Affiliations, email...

DOI : <http://dx.doi.org/10.1016/j.dib.2014.12.001>

Type de licence

Date de l'article : Date de soumission de l'article, date de publication, date de validation, date de révision.

Résumé :

Présentation du contexte d'obtention des données (front de recherche, question de recherche)

TABLE SPÉCIFIQUE

Domaine de recherche	Physique, chimie, psychologie, etc.
Domaine plus spécifique	Ex : physique nucléaire
Type de données	Tableau, image (radiographie, microscopie...), texte, graphique, figure, etc.
Moyens d'acquisition des données	Microscopie, enquête (vue générale), MEB, RMN, spectroscopie de masse, etc.
Format des données	Brut, filtré, analysé, etc.
Facteurs expérimentaux	Description brève de la préparation des échantillons
Caractéristiques expérimentales	Description expérimentale très brève
Emplacement de source de données	Ville, Pays, coordonnées GPS pour échantillons ou données
Accès aux données	Nom de l'entrepôt, son DOI, ou l'Url directe aux données
Article de recherche lié	Si les données accompagnent un article de recherche citez-le

Intérêt du jeu de données : Décrire la valeur scientifique de ces données

Description des données : Décrire brièvement les données partagées

Matériels et méthodes : Description complète de la méthode d'obtention des données. Inclure n'importe quelles figures/tables pertinentes à la compréhension des données.

Remerciements

Références : Références des données citées.



* Facultatif si données intégrées dans l'article

La minute "Publier un Data paper"

1 Choisir une revue



2 Choisir un entrepôt & déposer les données



3 Rédiger le data paper & la cover letter



4 Lier le data paper aux données déposées



5 Soumettre le data paper



6 Attribuer un identifiant au data paper





Contenus développés en partenariat avec l'Inra

Trois stratégies pour publier des données

- Intégrer les données dans un article publié dans une revue ;
- Joindre les données dans un fichier joint en tant que matériel supplémentaire ;
- Publier les données dans un data paper.

Article scientifique Matériel supplémentaire Data paper

Lorsqu'un article est publié dans une revue, les données qui y sont intégrées le sont aussi.



Avantages et limites de ce mode de publication de données :

- +** – Intégration maximale des données et de l'article : citable, recherché ;
- Paternité des données / crédits aux auteurs.
- – Données difficiles à trouver indépendamment de l'article et dans une forme peu ou pas réutilisable.

★★★★☆ Recherche et citabilité

Les données sont difficiles à trouver indépendamment de l'article dans lequel elles apparaissent.

★★★★★ Paternité et crédits auteurs

La reconnaissance de la paternité et l'accréditation aux auteurs sont possibles, comme pour un article classique.

★★★★☆ Volumétrie

Le volume de données pouvant être intégré est très contraint par les règles de rédaction de l'article.

★★★★☆ Réutilisabilité

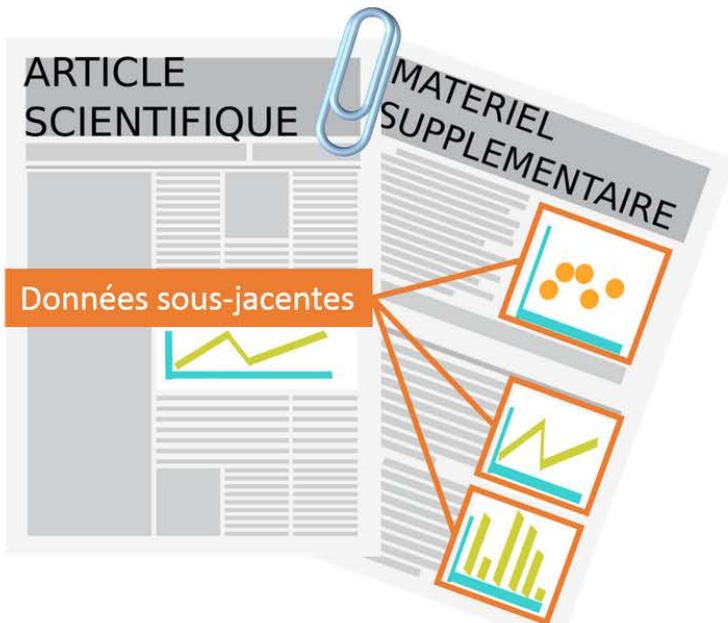
La forme dans laquelle les données sont publiées ne permettent pas ou peu leur réutilisation.

Trois stratégies pour publier des données

- Intégrer les données dans un article publié dans une revue ;
- Joindre les données dans un fichier joint en tant que matériel supplémentaire ;
- Publier les données dans un data paper.

Article scientifique Matériel supplémentaire Data paper

Les données sous-jacentes d'un article peuvent être jointes dans un fichier en tant que matériel supplémentaire. Ces données, qui vont plus loin que les données intégrées dans l'article, sont de plus en plus demandées aux auteurs pour des questions de validation.



Avantages et limites de ce mode de publication de données :

- +** – Format des données libéré des contraintes de rédaction de l'article ;
– Paternité des données / crédits aux auteurs.
- – Taille souvent limitée ;
– Peu de standardisation sur le signalement des fichiers « supplémentaires » ;
– Identification des données indépendamment de l'article possible mais rare ;
– Données difficiles à trouver indépendamment de l'article et dans une forme peu ou pas réutilisable.

Données intégrées

Matériel supplémentaire

Data paper

★★★★☆ **Recherche et citabilité**

Les données sont difficiles à trouver indépendamment de l'article qu'elles accompagnent. Il est rarement possible d'identifier les données (via un identifiant pérenne tel que le DOI) indépendamment de l'article.

★★★★★ **Paternité et crédits auteurs**

La reconnaissance de la paternité et l'accréditation aux auteurs sont possibles, comme pour un article classique.

★★★★☆ **Volumétrie**

La taille des données est souvent limitée.

★★★★☆ **Réutilisabilité**

La forme dans laquelle les données sont publiées ne permettent pas ou peu leur réutilisation.



Trois stratégies pour publier des données

- Intégrer les données dans un article publié dans une revue ;
- Joindre les données dans un fichier joint en tant que matériel supplémentaire ;
- Publier les données dans un data paper.

Article scientifique Matériel supplémentaire Data paper

Les data papers (data articles / data descriptors) sont des articles qui ont pour but de rendre les données accessibles, interprétables et réutilisables. Ils fournissent une voie formalisée au partage des données. Les data papers peuvent paraître dans des revues classiques ou dans des revues spécialisées, appelées data journals.

Avantages et limites de ce mode de publication de données :

-  – Paternité des données / crédits aux auteurs ;
- Citation aisée ;
- Réutilisation des données facilitée ;
- Données normalisées, standardisées, conservées de façon pérenne ;
- Pas de restriction en volume ;
- Liens vers les données déposées réciproques et sécurisés.
-  – Interrogation sur la qualité du peer-review ? ;
- Données déposées de préférence dans un entrepôt, impliquant la recherche et le choix de l'entrepôt adapté.

Données intégrées

Matériel supplémentaire

Data paper

★★★★★ **Recherche et citabilité**

Les données étant publiées au même titre qu'un article classique, on peut aisément les retrouver et les citer. Un lien réciproque est établi entre l'article et le jeu de données (déposé dans un entrepôt). Pour en savoir plus, vous pouvez consulter le site [DataCite](#).

★★★★★ **Paternité et crédits auteurs**

La reconnaissance de la paternité et l'accréditation aux auteurs sont possibles, comme pour un article classique.

★★★★★ **Volumétrie**

Il n'y a pas de restriction en volume des données.

★★★★☆ **Réutilisabilité**

La réutilisation des données est facilitée par la description des données. D'autres critères, indépendants du data paper et liés à la gestion des données, ont un impact sur la réutilisation des données (ex : le choix d'un format ouvert).

Les Identifiants pérennes

En bref



La fiche synthétique

Fiche 3 min PDF



La minute « Identifiants pérennes »

Vidéo 2 min

Les identifiants pérennes garantissent un lien stable à la ressource en ligne. Ils font correspondre en permanence l'identité de la ressource à sa localisation sur le web.

Il existe 2 grands types, les identifiants pérennes objet (pour les publications et données) et contributeur (pour les auteurs et institutions).


Consultez les sujets développés pour cette thématique et n'hésitez pas à laisser vos commentaires sur les pages des produits !

L'essentiel



Les identifiants pérennes : un aperçu

Guide 8 min



Petit guide DOI du chercheur

Infographie 2 min

Identifiants pérennes

Les identifiants pérennes sont des éléments importants du mouvement de l'Open Science. Il en existe 2 grands types : les **identifiants objet** pour vos productions scientifiques (publications et données) et les **identifiants contributeur** pour les auteurs et les institutions. On parle souvent de PID pour Persistent Identifier en anglais.

Utilité des PID

Identification et découverte

Les identifiants pérennes contributeur permettent l'identification fiable d'un auteur ou d'une institution. Les identifiants objet permettent l'identification fiable des productions scientifiques et concourent à leur découverte.

Notes

Le DOI (Crossref et DataCite) pour les ressources et l'ORCID pour les contributeurs sont les identifiants les plus utilisés : plus de 130 millions de DOI et plus de 3 millions d'ORCID au 01/03/2017.

Accès, partage et diffusion

Les identifiants pérennes favorisent et facilitent le partage, la réutilisation des productions scientifiques et permettent leur accès sur le long terme. Ils simplifient aussi leur citation.

L'ouverture des données permet d'éviter la redondance et d'économiser du temps et des ressources financières qui seront ainsi allouées à d'autres projets de recherche.

Le saviez-vous ?

Plusieurs systèmes d'identifications collaborent pour offrir des services permettant d'agrèger les productions scientifiques, c'est-à-dire les publications et les données, à leur(s) auteur(s) et leur(s) institution(s).
Ex : Auto-Update Service = Crossref (publications) + DataCite (données) + ORCID (collaborateurs).

Deux types de PID

Les identifiants objet

Les PID objet (pour les publications et les données) permettent également de faire le lien entre les articles publiés et les jeux de données sous-jacents.

Exemples

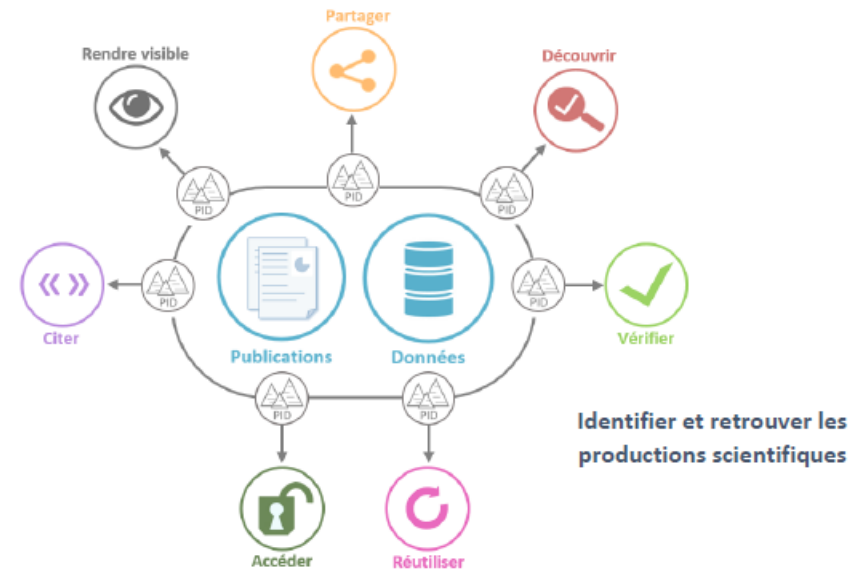
Handle, DOI (Identifiant numérique d'objet), ePIC (european Persistent Identifier Consortium), ARK (Archive Resource Key), PURL (persistent URL)...

Les identifiants contributeur

Les PID contributeur (pour les auteurs et institutions) visent aussi à désambigüiser les noms et résoudre les problèmes d'homonymie, translittération, etc. Ils augmentent ainsi la visibilité académique.

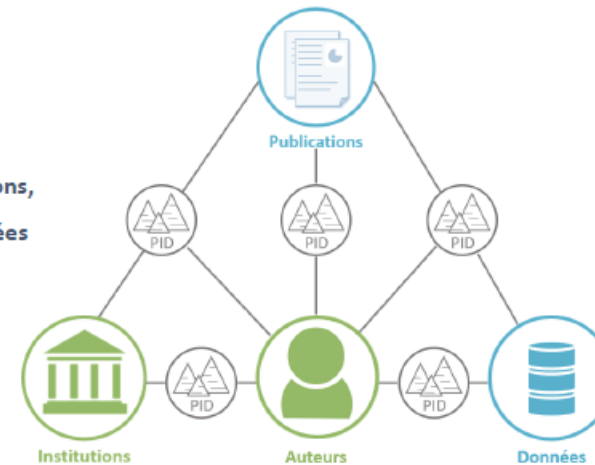
ORCID (Open Researcher and Contributor ID), identifiants du monde des bibliothèques (ISNI, IdRef...), d'éditeurs commerciaux, de réseaux sociaux, d'archives ouvertes (IdHAL, ArXiv Author ID)...

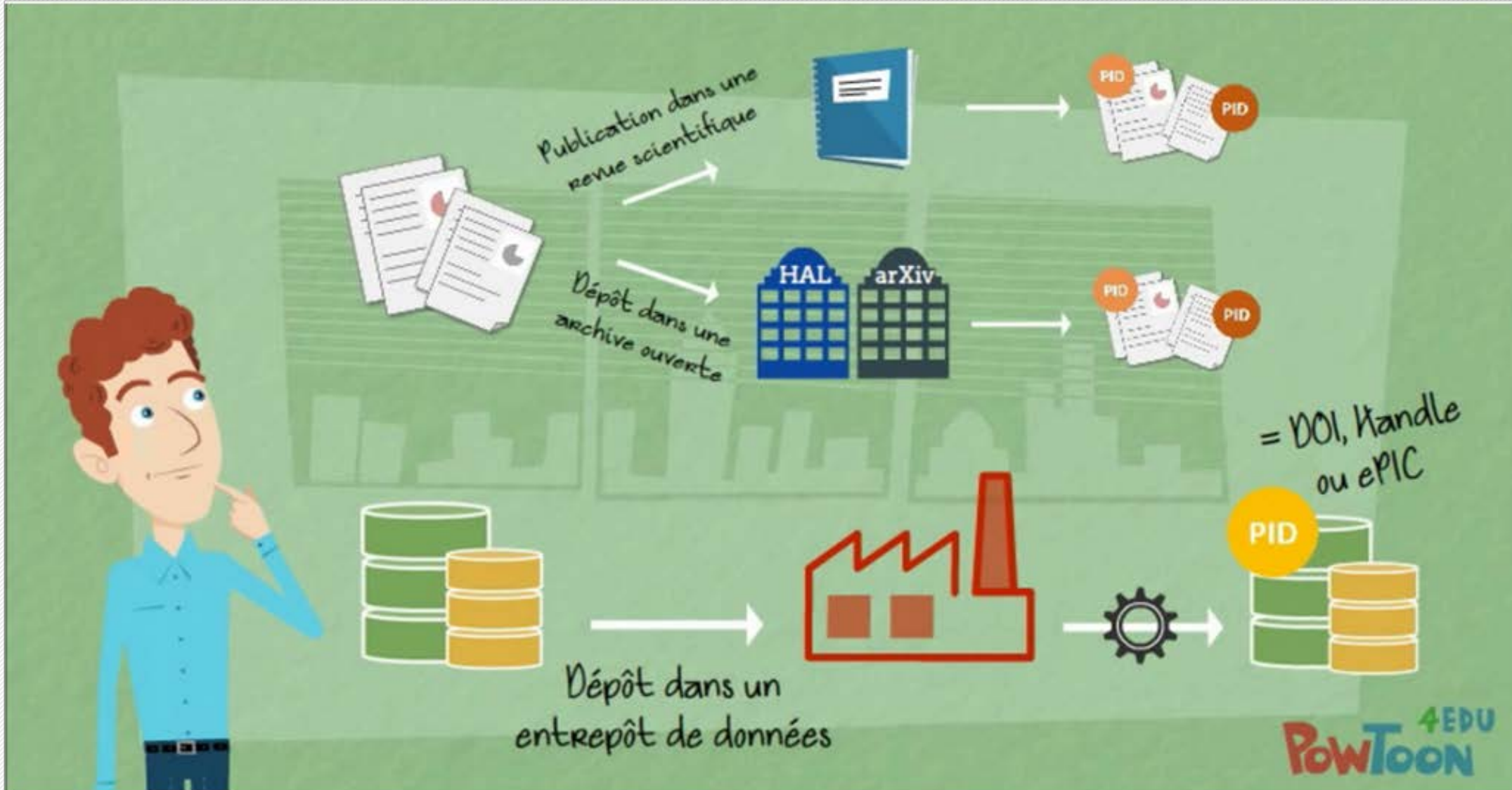
Identifiants pérennes



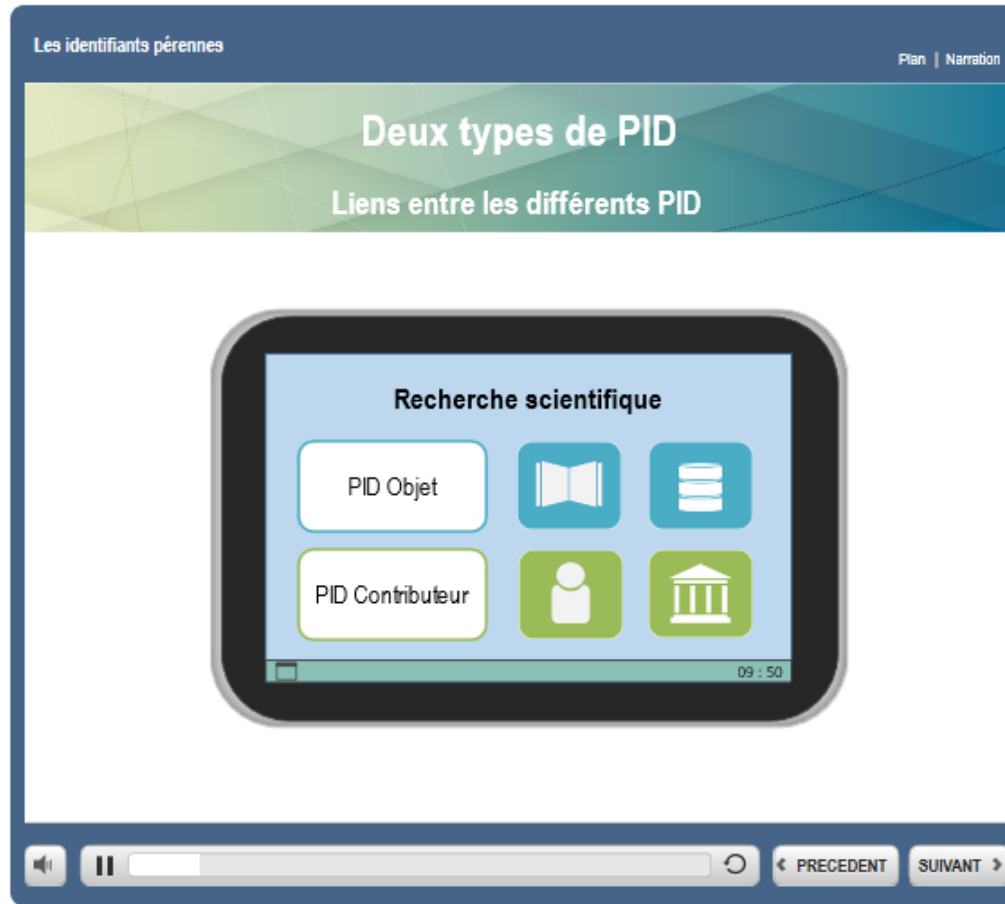
Identifier et retrouver les productions scientifiques

Lier auteurs, institutions, publications et données





Un aperçu







Les identifiants pérennes Plan | Narration

Deux types de PID

Liens entre les différents PID

Recherche scientifique

PID Objet		
PID Contributeur		

09 : 50

◀ PRECEDENT SUIVANT ▶

Le sondage DoRANum



Besoins, attentes et disponibilité

Structure de l'enquête



- **Le profil**
- **Les usages :**
 - ✓ gestion données recherche
 - ✓ apprentissage
- **Les besoins :**
 - ✓ gestion données recherche
- **Les attentes :**
 - ✓ modalités d'apprentissage

Diffusion de l'enquête



- **Questionnaires papiers :**
 - Lors de manifestations
 - Lors de formations
 - À la disposition des partenaires
- **Questionnaire en ligne :**
 - Via les réseaux des partenaires du projet
 - Sur le site web du projet (<http://www.dorandum.fr>)

L'enquête reste accessible à l'adresse:

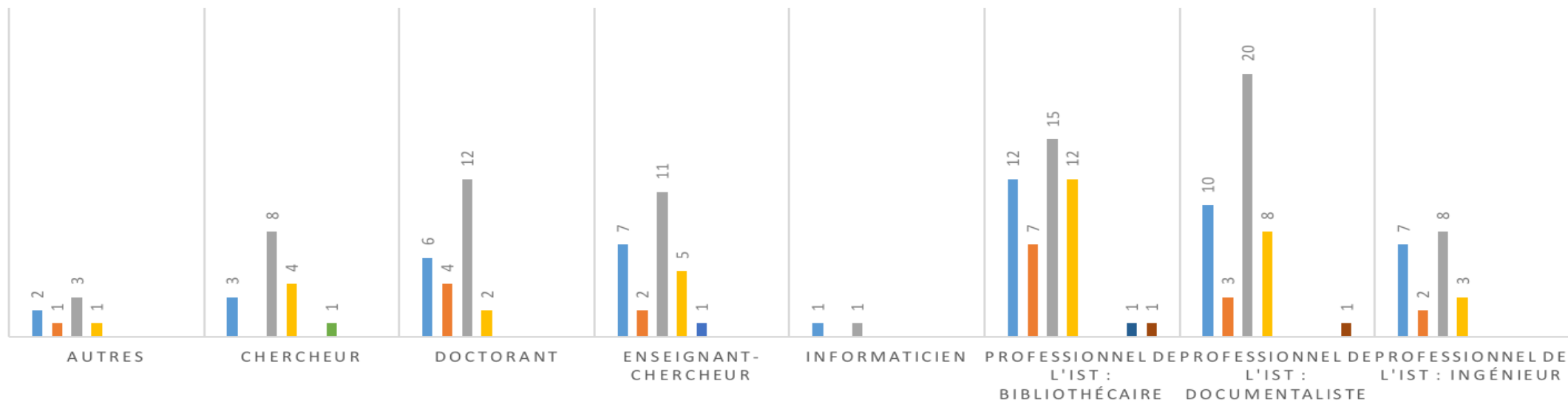
<https://framaforms.org/gestion-des-donnees-de-la-recherche-queelles-sont-vos-attentes-en-termes-dapprentissage-et-1479885542>

✓ **128 répondants** au 28/10/2016 (**208** au 15/02/2017)

Ressources d'apprentissage

LES ATTENTES EN TERMES DE RESSOURCES D'APPRENTISSAGE SUR LE THÈME DE LA GESTION DES DONNÉES DE LA RECHERCHE

- Des ressources audiovisuelles (didacticiels)
- Des ressources audiovisuelles (interviews d'experts)
- Des ressources écrites : fiches synthétiques
- Des ressources écrites : guides
- Neant
- Ressources écrites ET audiovisuels
- toutes
- exemple en ligne



Merci de votre attention !



<http://www.dorandum.fr>

Participez à l'enquête !



Contacts:

Paolo Lai paolo.lai@inist.fr & Yvette Lafosse yvette.lafosse@inist.fr

Michel Roland michel.roland@unice.fr