




1 Valorisation des données

2 points

Pour valoriser mes données, je peux :

- VRAI**
- les exposer sur un site internet
 - les cartographier
 - rédiger un data paper
- FAUX** - les archiver définitivement

En complément du dépôt dans un entrepôt, il peut s'avérer utile, surtout dans le cas de données nombreuses et complexes, de les exposer sous forme visuelle (cartographies, graphiques, etc.) via une plateforme. L'outil Omeka est un logiciel libre couramment utilisé, en particulier sa





version sémantique appelée Omeka S.

Exemple : le site CoReA, réalisé grâce à l'outil Omeka, permet de naviguer dans des corpus et ressources archéologiques du centre Camille Jullian.

En SHS, l'IR* Huma-Num propose l'entrepôt Nakala qui comprend un système de publication (NAKALA-PRESS) et permet ainsi d'exposer les données déposées.

Le data paper est une publication centrée sur les données de recherche. Il permet de présenter le projet dans sa totalité en facilitant l'accès aux jeux de données.

Vous pouvez consulter les ressources DoRANum sur les thématiques :



- **[Accès et visualisation](#)**
- **[Data papers et Data journals](#)**





2 Rédaction du DMP ou PGD, plan de gestion de données

2 points

La rédaction d'un DMP ou PGD intervient dès le début du projet et figure parmi les livrables demandés par les financeurs.

La première version du PGD peut être demandée dès le dépôt du projet en même temps que les demandes de financements. Il s'agira à cette étape de décrire, dans les grandes lignes, les choix envisagés.

Le PGD est un document évolutif, il sera donc complété au fur et à mesure de l'avancée du projet.





Il doit aider à organiser et anticiper toutes les étapes du cycle de vie de la donnée, et ce, pour chaque jeu de données, depuis la création des données jusqu'à leur partage et leur archivage.

De nombreux modèles de PGD ont été établis par des organismes, instituts, financeurs à destination de leurs utilisateurs.

Il existe également des outils dédiés à la rédaction du DMP comme DMP OPIDoR.

Vous pouvez consulter :



- les ressources sur le site [DoRANum](#), rubrique "Plan de gestion de données"
- le site [DMP OPIDoR](#)



3 Publication d'un data paper

3 points

Le data paper est une publication centrée sur les données de recherche. Il permet de décrire les données et les métadonnées associées de manière plus approfondie que dans une publication classique.

Un data paper détaille les processus de création, collecte, traitement et analyse des données.

Certaines revues proposent des data papers en complément des articles. Il existe aussi des data journals qui ne contiennent que des data papers.

Exemple : le data journal gratuit RIO (Research Ideas and Outcomes).





Le data paper permet au chercheur de présenter le projet dans sa totalité en facilitant l'accès aux jeux de données, ce qui accroît leur réutilisation et leur citabilité.



Vous pouvez faire appel à un chercheur du domaine



Vous pouvez consulter les ressources de la rubrique "[Data papers et Data journals](#)" sur le site DoRANum





4

Licence attribuée aux données publiques

3 points

Exemples de licence adaptée aux données publiques :

- **Licences Creative Commons (ex: CC-BY)**
- **Licence ouverte (Etalab)**

Le fait de choisir une licence de diffusion permet d'afficher clairement les modalités de réutilisation.



Vous pouvez faire appel à un juriste/service juridique





Vous pouvez consulter :



- les ressources DoRANum, rubrique "Aspects juridiques, éthiques, intégrité scientifique"
- la liste des licences de réutilisation établie dans le cadre de la loi pour une République numérique
- l'outil "Licence Selector"
- l'outil "Choose an open source licence"





5

Règles de nommage des fichiers de données

I point

Voici quelques éléments à prendre en compte pour nommer un fichier de données :

- Commencer par une lettre
- Utiliser 30 caractères maximum
- Pas de caractères spéciaux, accents, espace ou point
- Utiliser les lettres (majuscules et minuscules), chiffres, tirets, numéro ou acronyme du projet
- Date de création sous la forme yyyy-mm-dd ou yyyymmdd





- Nom du créateur
- Description brève du contenu
- Numéro de version
- Extension du format

Exemple :

Relevés de transitions de végétation (tv) effectuées
par Andrew D. Richardson (ADR) dans le cadre du
projet Phenocam (Ph) sous un format csv :
Ph_2012-08-29_ADR_tv_VI.csv



Vous pouvez consulter la
ressource "[Comment bien
nommer ses fichiers](#)" sur le
site DoRANum





6

Format de fichier de données

2 points

Ce sont des formats ouverts :

VRAI - .csv **FAUX** - .doc
- .pdf - .avi
- .txt

Recommandations :

- Pour faciliter la réutilisation et la conservation des données, utilisez des formats non propriétaires ou largement utilisés afin de prévenir les potentiels problèmes d'obsolescence

Exemples de formats ouverts, standardisés et pérennes : .csv, .txt, .png, .pdf, .mp4...





- La conservation des versions de logiciels utilisés doit être envisagée s'il s'agit de formats propriétaires et si les données doivent être reproduites dans le temps
- Prévoyez du temps pour convertir les fichiers
- Vérifiez la compatibilité du format avec l'outil FACILE du CINES

Vous pouvez consulter :



- le site du CINES
- la ressource [Format ouvert ou fermé ?](#) sur DoRANum





7

Choisir un entrepôt de données

2 points

Il est recommandé de choisir prioritairement un entrepôt disciplinaire et si aucun ne convient, de déposer dans l'entrepôt de votre institution, s'il en existe un.

Si aucun entrepôt disciplinaire ou institutionnel ne correspond à vos besoins, vous pouvez déposer vos données dans l'entrepôt national pluridisciplinaire

[Recherche Data Gouv.](#)

Si vous pouvez, privilégiez un entrepôt certifié et/ou « de confiance », en accès ouvert.



À noter : en lien avec l'archive ouverte HAL, Software Heritage assure l'accessibilité et la conservation des codes sources en libre accès tout en gérant les versions.



Pour choisir le bon entrepôt, appuyez-vous sur les pratiques de votre communauté scientifique !

Vous pouvez consulter :



- les ressources de la rubrique ["Dépôt et entrepôts"](#) sur le site DoRANum
- l'outil [Cat OPIDoR](#)
- le répertoire d'entrepôts [re3data](#)
- [Open Access Directory](#)
- [Software Heritage](#) pour le dépôt des codes sources



8

Stockage des jeux de données

I point

Il est important de sauvegarder ses données tout au long du projet.

Il est recommandé d'utiliser un espace collaboratif de stockage sécurisé ou un serveur sécurisé pour éviter la perte de données.

Exemple :

L'espace collaboratif de travail du CNRS (CoRe) permet :

- le stockage et la sauvegarde des données
- la conservation des versions successives des fichiers modifiés
- l'accessibilité à distance 24 h/24 et 7 jours/7
- l'accès contrôlé





**Vous pouvez faire appel à un
informaticien spécialisé**



9 Autorisations légales

2 points

Toutes vos données ne sont peut-être pas diffusables.

- **Y a-t-il des données à ne pas divulguer ?**
- **Y a-t-il des données personnelles ?**

Si oui : comment se fera l'anonymisation ?

- **Y a-t-il des restrictions ou un délai à respecter pour l'accès aux données ?**

Exemples : diffusion d'images, d'interviews, d'enregistrements.



**Vous pouvez faire appel au délégué
à la protection des données ainsi
qu'à un juriste/service juridique**





Vous pouvez consulter :



- La ressource DoRANum sur
"La communicabilité des données
de la recherche"

- le poster "Qui a les droits, quelles obligations ?"
de l'Ecole des Ponts ParisTech

- La liste des Organismes ayant désigné un(e)
délégué(e) à la protection des données (DPD/DPO)
sur data.gouv.fr





10 Préparation du dépôt des données dans un entrepôt

3 points

Déposer vos données dans un entrepôt dédié permet de les partager.

Pour préparer ce dépôt, il faut :

- sélectionner les données et métadonnées à déposer
- enrichir les métadonnées en fonction du standard proposé par l'entrepôt ou du standard choisi
- vérifier la compatibilité des formats de données. Si besoin, migrer les données vers un format adapté
- préparer si nécessaire les codes sources qui permettront de lire et traiter les données.





Vous pouvez faire appel à un documentaliste, à un informaticien.

Vous pouvez consulter :



- la ressource DoRANum
"Les standards de métadonnées : pourquoi et lequel ?"

- la ressource DoRANum "Format ouvert ou fermé ?"

- la ressource DoRANum "Vérifier ses données de recherche"



11

Licence ouverte

2 points

La licence CC BY est l'équivalent international de la licence ouverte française Etalab. C'est l'une des licences les plus ouvertes. La seule condition est de citer l'auteur.

La CC BY NC SA est beaucoup plus restrictive qu'Etalab car elle impose de mentionner la paternité, elle empêche toute utilisation commerciale et oblige au partage dans les mêmes conditions.

La CCO est la plus ouverte et n'impose aucune restriction.



À noter qu'en France, le droit de paternité (citation du nom de l'auteur) doit être respecté. Il s'agit d'un droit moral, inaliénable, perpétuel et incessible. L'utilisation de la licence CC0 n'est donc pas recommandée. Si l'entrepôt de données choisi l'impose, il faudra veiller à ajouter le nom de auteurs dans les métadonnées.



Vous pouvez faire appel à un juriste ou service juridique pour vous aider.

Vous pouvez consulter :



- la ressource DoRANum sur les "[licences de réutilisation dans le cadre de l'open data](#)"

- la liste des [licences de réutilisation](#) établie dans le cadre de la loi pour une République numérique, sur le site data.gouv.fr



12

Identifiant AUTEUR unique et pérenne

3 points

Ce sont des identifiants AUTEUR :

VRAI

FAUX

- ORCID

- DOI

- IdHAL

- ARK

Il est fortement recommandé d'avoir un
identifiant auteur pour partager ses
publications, données et logiciels.

Cela permet :

- d'éviter les confusions liées à l'homonymie
- d'associer les différentes identités
du chercheur (nom avec une ou





plusieurs initiales de prénoms, changement de nom pour les femmes)

- d'éviter au chercheur de ressaisir les informations le concernant

Un identifiant auteur constitue ainsi une clé unique qui facilite la recherche d'information.

Recommandation : ORCID



Vous pouvez consulter les ressources pédagogiques de DoRANum sur les [identifiants pérennes](#)






13 Logiciel(s) pour accéder et lire les données

3 points

Si un logiciel spécifique est nécessaire, il faut : **FAUX** - Garder le format d'origine
VRAI - Fournir le nom du logiciel
- Fournir le script

Dans le cas où un logiciel est nécessaire pour accéder et lire vos données :

- soit vous devrez fournir le logiciel et si besoin le script associé,
- soit vous devrez convertir vos données dans un format ouvert si vous avez utilisé un logiciel propriétaire





Exemple :

**Vous avez utilisé le logiciel propriétaire Excel,
vos données sont donc au format .xls.**

**Pour permettre leur accès et leur partage,
vous devrez les convertir au format .csv.**

Recommandations :

- utiliser un logiciel Open source de
préférence**
- rendre accessible la documentation sur
le logiciel**



**Vous pouvez faire appel à un
informaticien spécialisé**





14 Métadonnées embarquées et métadonnées enrichies

2 points

Les métadonnées embarquées existent pour un grand nombre d'objets numériques, et sont créées automatiquement.

Par exemple :

- dans les photos : marque de l'appareil, localisation
- dans les documents texte : date, auteur, taille...

Ces métadonnées "automatiques" doivent être complétées par des métadonnées qui permettent une description plus précise de vos données : ce sont les métadonnées enrichies.

Il est nécessaire de les renseigner lors du dépôt dans un entrepôt.





Recommandations :

- Décrire les métadonnées au fur et à mesure de l'avancée du projet
- Associer un fichier de métadonnées à chaque fichier de données



Vous pouvez faire appel à
un documentaliste pour vous aider

Vous pouvez consulter :



- les ressources DoRANum
sur les métadonnées,
standards, formats
- le "Guide de bonnes pratiques" sur le site ORDaR





15

Données sensibles

2 points

Ce sont des données sensibles :

VRAI - appartenance ethnique d'une famille dans une étude ethnologique
- opinions politiques des personnes dans une interview

FAUX - géolocalisation d'une espèce végétale protégée

Les données sensibles sont des données à caractère personnel qui font apparaître, directement ou indirectement :

- l'origine raciale ou ethnique





- les opinions politiques
- les convictions religieuses ou philosophiques
- l'appartenance syndicale
- les informations génétiques ou biométriques.

L'exemple de l'espèce végétale protégée n'est pas une donnée sensible mais il nécessite une vigilance particulière quant à sa communicabilité.



Vous pouvez faire appel à un juriste/service juridique





16

Standard de métadonnées généraliste

3 points

Un standard de métadonnées est un modèle permettant de décrire les données de manière plus précise.

Le standard est choisi en fonction de la destination des données (dépôt, publication, archivage, etc.). Il peut aussi être spécialisé par discipline, par type de données...

Si aucun standard spécifique ne répond à vos besoins, vous pouvez utiliser un standard de métadonnées généraliste, comme le Dublin Core.



Les 15 éléments du Dublin Core :

Élément	Élément (anglais)
1. Titre (métadonnée)	Title
2. Créateur (métadonnée)	Creator
3. Sujet (métadonnée) ou mots clés	Subject
4. Description (métadonnée)	Description
5. Éditeur	Publisher
6. Contributeur	Contributor
7. Date (métadonnée)	Date
8. Type	Type
9. Format	Format
10. Identifiant de la ressource	Identifier
11. Source	Source
12. Langue (métadonnée)	Language
13. Relation (métadonnée)	Relation
14. Couverture (métadonnée)	Coverage
15. Gestion de droits (métadonnée)	Rights



**Vous pouvez faire appel à un
documentaliste ou un informaticien**

Vous pouvez consulter :

- la ressource "[Les standards de métadonnées : pourquoi et lequel ?](#)" sur le site DoRANum

- "[Disciplinary Metadata](#)" sur le site du DCC



- "[Metadata Standards Catalog](#)" de la
Research Data Alliance (RDA)

- le site [FAIRsharing.org](#)



17

Vocabulaire de description des données

2 points

**C'est un vocabulaire contrôlé/standard,
par exemple un lexique, un thésaurus,
une ontologie...**

Il existe 2 types de vocabulaires :

- le vocabulaire contrôlé (thésaurus)**
- le vocabulaire libre (mots-clés)**

**Utiliser un vocabulaire contrôlé facilite
l'intégration des données et leur partage
avec votre communauté scientifique.**



Exemples de thésaurus des bases de données en sciences de l'éducation :

Thésaurus
ERIC

PsycINFO

Thésaurus
Education
source



Vous pouvez faire appel à un
documentaliste



Vous pouvez consulter le "[Guide de
bonnes pratiques](#)" sur le site ORDaR



18

Identifiant unique et pérenne pour les données

2 points

Un identifiant unique et pérenne facilite le suivi, la localisation, l'accès et la citation des données lors de leur publication ou à des fins de réutilisation.

Un identifiant est souvent proposé par l'entrepôt choisi.

Recommandations :

Privilégiez un identifiant DOI. Si l'URL liée à vos données a été modifiée, il est indispensable de la mettre à jour.





Vous pouvez faire appel à l'Inist-CNRS

Vous pouvez consulter :



- les ressources de DoRANum sur les identifiants pérennes
- le site OPIDoR



19 Archivage des données

I point

L'archivage numérique pérenne consiste à conserver les fichiers de données, les logiciels et leurs métadonnées à moyen et long terme.

Cela nécessite quelques préparatifs :

- Sélectionner les données et métadonnées à conserver
- Évaluer la volumétrie
- Planifier l'archivage
- Envisager les coûts de cette opération

L'organisme principal chargé de l'archivage en France est le CINES.



À savoir : certains entrepôts proposent aussi un archivage à long terme.

Par exemple : 4TU.Centre for Research Data.

Vous pouvez faire appel à :



- un archiviste orienté IST (information scientifique et technique)
- au CINES,
- à Huma-Num (en SHS)
- à un informaticien spécialisé

Vous pouvez consulter :



- les ressources DoRANum sur le [stockage et l'archivage](#)
- le site [Software Heritage](#)
- le site du CINES sur le [Le concept d'archivage numérique pérenne](#)



20

DMP ou Plan de gestion de données

I point

DMP signifie Data Management Plan.

Un DMP (ou PGD, pour Plan de gestion de données), est un document explicitant la façon dont le chercheur obtient, documente, analyse et utilise ses données à la fois au cours de sa recherche et une fois le projet terminé.

Il décrit dans le détail les méthodes et processus de création, de fourniture, de maintenance, de conservation et de protection des données.







21

Dépôt des données dans un entrepôt

3 points

Déposer ses données dans un entrepôt permet :

- la valorisation de vos travaux de recherche en facilitant la validation par les pairs et la citation
- l'émergence de nouvelles recherches
- la diminution des coûts de production quand les recherches sont trop chères à dupliquer ou pour des recherches impossibles à reproduire car elles ont généré des données uniques (par ex : observation d'une supernova, interview de poilus, etc.)





22 Données personnelles

3 points

Pour m'aider, je peux faire appel :

VRAI - Au Délégué à la Protection des Données

- À un juriste

FAUX - Au directeur de l'unité

- À un informaticien

- À un documentaliste

L'aide que peut m'apporter le Délégué à la Protection des Données :

- Informations et conseils en amont de tout projet impliquant le traitement de données personnelles



- Constitution et mise à jour du registre des traitements mis en œuvre conformément à la loi

- Protection des droits des personnes concernées par les traitements et réponse aux demandes de droits d'accès, d'opposition, de modification et de suppression

Les sujets sur lesquels le juriste/service juridique peut m'apporter une aide :

- Propriété intellectuelle des données
- Communicabilité des données
- Protection des données
- Choix des licences

Vous pouvez consulter :



- les ressources DoRANum sur la thématique Aspects juridiques, éthiques, intégrité scientifique

- La liste des Organismes ayant désigné un(e) délégué(e) à la protection des données (DPD/DPO) sur data.gouv.fr



23 Rédaction du DMP ou Plan de gestion de données


I point

Les personnes qui peuvent m'aider à rédiger un Plan de gestion de données :

VRAI - un documentaliste
- un informaticien
- le Délégué à la Protection des Données

FAUX - un gestionnaire du laboratoire

Le Plan de Gestion de Données (PGD ou DMP) est un document évolutif qui explicite la façon dont le chercheur obtient, documente, analyse et utilise ses





données à la fois au cours de sa recherche et une fois le projet terminé.

Il décrit en détail les méthodes et processus de création, de fourniture, de maintenance, de conservation et de protection des données.

Le chercheur est donc la personne la mieux placée pour décrire les jeux de données et leur contexte.

Cependant, il peut être aidé pour certaines parties du PGD plus techniques (métadonnées, standards, formats, archivage, etc.).

Il ne faut pas hésiter à solliciter les personnes ressources de son institut : documentaliste, informaticien, juriste... ainsi que le Délégué à la Protection de Données.

Pour le trouver, vous pouvez consulter la liste des Organismes ayant désigné un(e) délégué(e) à la protection des données (DPD/DPO) sur data.gouv.fr





24

Inist-CNRS

2 points

L'Inist-CNRS propose les services suivants :

- VRAI** - Attribution de DOI
- Exposition des données
- Aide à la rédaction d'un DMP
- FAUX** - Stockage des données

L'Inist-CNRS propose de nombreux services d'accompagnement au chercheur :

- Accompagnement à travers l'élaboration et la mise en pratique de plans de gestion de données et de logiciels





- DMP OPIDoR : intégration de modèles de rédaction de DMP institutionnels
- PID OPIDoR : attribution de DOI
- Cat OPIDoR : catalogue d'outils et services
- Exposition des données via Omeka
- DoRANUM : Accompagnement et formation sur les données de la recherche





25

Informaticien

3 points

Un informaticien peut m'aider dans les cas suivants :

VRAI - Choix des formats de fichiers

- Sécurité des données

- Stockage des données

FAUX - Renseignement des métadonnées

- Détermination de la propriété des données

Le chercheur est le mieux placé pour traiter et renseigner ses données mais il ne faut pas hésiter à faire





appel à un informaticien dans les opérations plus techniques telles que :

- le traitement informatique des données**
- la sécurité des données**
- le choix et la migration des formats de fichiers**
- le choix des standards de métadonnées**
- le dépôt dans un entrepôt**
- le stockage**
- l'archivage**





26 Données éligibles à l'archivage pérenne

3 points

Les données de recherche éligibles à l'archivage pérenne sont :

VRAI :

- les données coûteuses à produire
- les données non-reproductibles
- les données à valeur patrimoniale

FAUX :

- toutes les données de recherche
- les données dans un format de fichier rare

Les données de recherche ne peuvent pas toutes être éligibles à l'archivage.





Il y a tout un ensemble de critères à prendre en compte et respecter :

- la valeur patrimoniale/scientifique des données
- le caractère unique/non reproductible des données
- le coût très élevé de production de ces données, l'archivage étant une opération également coûteuse
- la possibilité de maintenir le format de fichier dans un temps long
- ...

Dans tous les cas, cette décision est du ressort de la tutelle du chercheur et nécessite de se faire accompagner par un archiviste.

Vous pouvez consulter :



- les ressources DoRANum sur le [stockage et l'archivage](#)
- le site du [CINES](#)





27

Licence attribuée aux logiciels

3 points

Il existe plus de 300 licences pour les logiciels. Les plus courantes sont Apache, BSD, CeCILL, GNU et MIT.

Le fait de choisir une licence de diffusion permet d'afficher clairement les modalités de réutilisation du logiciel.



**Vous pouvez faire appel à un
juriste/service juridique**





Vous pouvez consulter :



- la ressource DoRANum sur les "[licences de réutilisation dans le cadre de l'open data](#)"
- la ressource DoRANum "[guide des licences ouvertes](#)"
- la liste des [licences de réutilisation](#) établie dans le cadre de la loi pour une République numérique sur le site data.gouv.fr
- l'outil [Licence Selector](#)
- l'outil [Choose an open source licence](#)





28

Licence attribuée aux bases de données

3 points

Il existe une licence spécifique aux bases de données : la licence ODbL. Créée dans le cadre de l'Open Data Commons, elle est utilisée aussi bien en France qu'à l'étranger. Les licences Etalab et Creative Commons peuvent aussi être appliquées.

En France, les deux licences exclusives que peut employer un acteur public pour ses données en libre accès sont ODbL et Etalab.

Le fait de choisir une licence de diffusion permet d'afficher clairement les modalités de réutilisation de la base de données.





Vous pouvez faire appel à un juriste/service juridique



Vous pouvez consulter :

- la ressource DoRANum sur les "[licences de réutilisation dans le cadre de l'open data](#)"
- la ressource DoRANum "[guide des licences ouvertes](#)"
- la liste des [licences de réutilisation](#) établie dans le cadre de la loi pour une République numérique, sur le site data.gouv.fr
- l'outil [Licence Selector](#)
- l'outil [Choose an open source licence](#)
- la ressource DoRANum sur les "[aspects juridiques des bases de données](#)"



29 Standards de métadonnées par discipline

3 points

Un standard de métadonnées est un modèle permettant de décrire les données de manière précise.

Il facilite la recherche et l'interopérabilité entre différents systèmes.

Il peut exister des standards spécifiques à votre discipline. Quand c'est le cas, il est recommandé d'y avoir recours.

Exemples de standards disciplinaires :

- LOM (Learning Object Metadata) en éducation



- 
- EML (Ecological Metadata Language) en écologie
 - DDI (Data Documentation Initiative) pour les statistiques et données d'enquêtes en Sciences humaines et sociales



Vous pouvez faire appel à un documentaliste

Vous pouvez consulter :

- la ressource "[Les standards de métadonnées : pourquoi et lequel ?](#)" sur le site DoRANum

- "[Metadata Standards Catalog](#)" de la Research Data Alliance (RDA)



- "[Disciplinary Metadata](#)" sur le site du DCC

- le site [FAIRsharing.org](#)

30

DMP ou plan de gestion de données

I point

PGD signifie Plan de Gestion de Données.

Un PGD (ou DMP, pour Data Management Plan), est un document explicitant la façon dont le chercheur obtient, documente, analyse et utilise ses données à la fois au cours de sa recherche et une fois le projet terminé.

Il décrit de manière précise les méthodes et processus de création, de fourniture, de maintenance, de conservation et de protection des données.



**Le PGD explique également la manière
dont les données seront partagées et
archivées sur le long terme.**





31 Rédaction du DMP ou Plan de gestion de données

I point

FAUX

Dans la première version de votre DMP, vous n'êtes pas obligé de répondre à toutes les questions. En effet, vous n'avez probablement pas toutes les réponses à ce stade du projet.

Cependant, il est important de s'appuyer sur les questions du DMP pour anticiper les problématiques d'ordre technique (stockage, partage, etc...) ou juridique afin de vous assurer une gestion efficace de vos données tout au long du projet et quand il sera terminé.





32 Entrepôts de données

2 points

Voici quelques exemples d'entrepôts de données généralistes :

- Recherche Data gouv
- Zenodo
- Figshare





33 Entrepôts de données disciplinaires

3 points

Voici quelques exemples d'entrepôts de données disciplinaires :

- **NAKALA** : entrepôt national français en SHS
- **SEANOE** : entrepôt national français en sciences marines
- **Quetelet-Progedo Diffusion** : entrepôt national français en sciences sociales
- **ORTOLANG** : entrepôt national français en sciences du langage
- **PANGAEA** : entrepôt international certifié en sciences de la Terre et de l'environnement
- **ArkeOpen** : entrepôt en archéologie

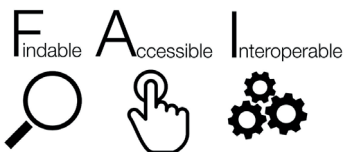
- GBIF : entrepôt international certifié sur la Biodiversité
- TreeBASE : données sous-jacentes aux publications en phylogénétique

Pour trouver l'entrepôt de données disciplinaire qui vous convient, vous pouvez consulter :

- la liste des entrepôts de confiance sélectionnés par le Collège Données de la recherche sur le site Ouvrir la Science



- Cat OPIDoR, rubrique "Entrepôt de données"
- les exemples d'entrepôts de données en SHS sur le site DoRANum
- la ressource sur les critères pour choisir un entrepôt de données sur le site DoRANum

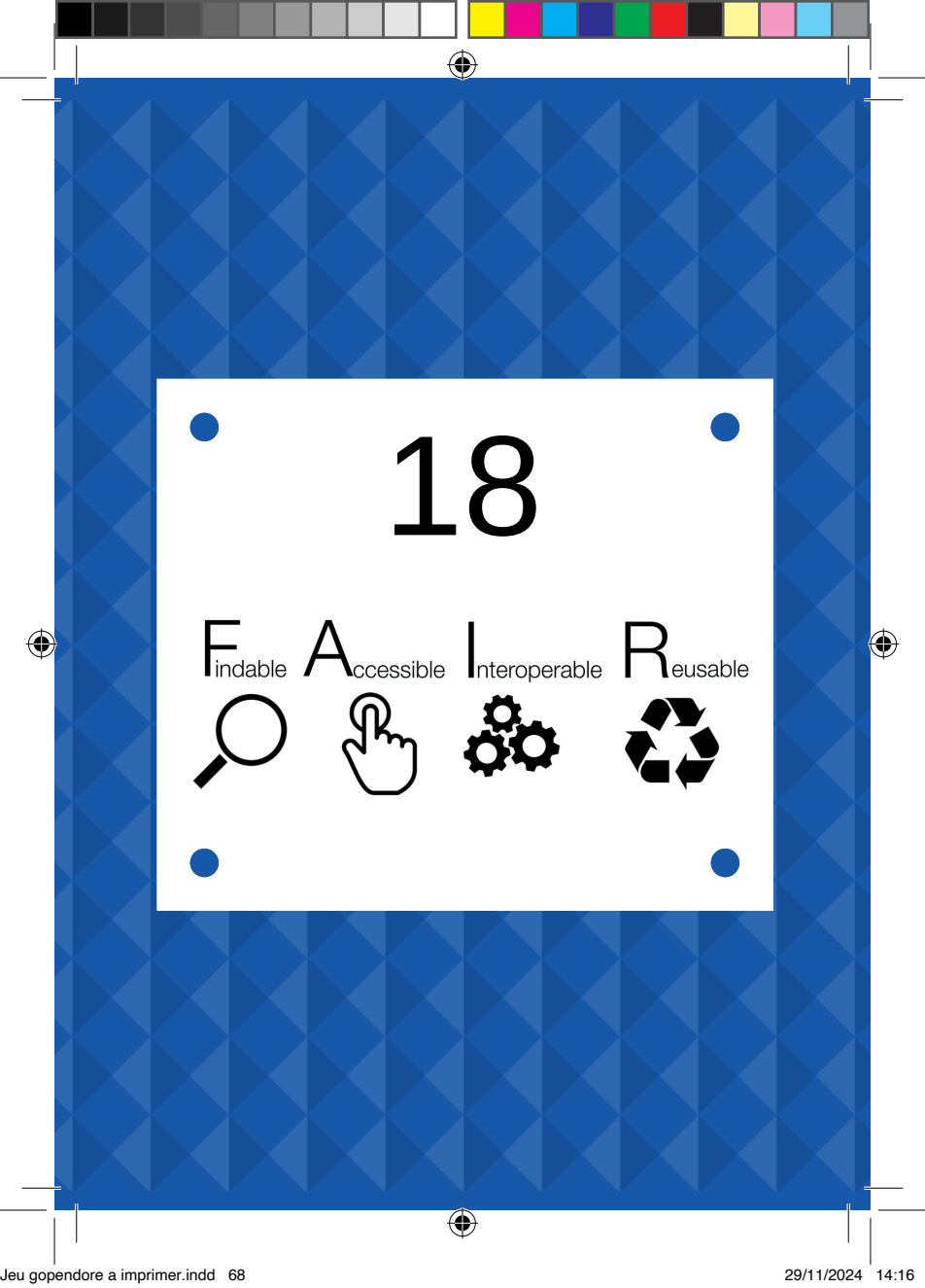


Où partager mes données pour permettre leur réutilisation ? (2 bonnes réponses)

- Sur un réseau social comme Academia ou ResearchGate
- Dans un entrepôt thématique de confiance
- Dans l'entrepôt national Recherche Data Gouv
- Dans une base de données disciplinaire accessible sur le site web de mon laboratoire

3 points





18

F_{indable}



A_{ccessible}

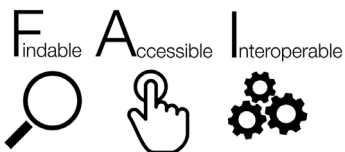


I_{nteroperable}



R_{eusable}





**Un data paper facilite
l'accès aux données.**

Vrai ou Faux ?

2 points





20

Findable Accessible Interoperable Reusable





3

Où trouver facilement des données de recherche ?

Bonnes réponses

- dans les data papers
- dans les entrepôts de données
- dans un moteur de recherche dédié aux données

À nuancer

- dans les "supplementary data"
- dans HAL

- Les "**supplementary data**" sont une solution mise en place par les éditeurs : ce sont des publications complémentaires sur les données, en lien avec un article scientifique.

La possibilité d'exploitation des données





est toutefois bien moindre que dans un entrepôt : formats de fichier souvent non-exploitable, revues sur abonnement...

- La mission principale de l'**archive ouverte nationale HAL** est le dépôt et la diffusion d'articles scientifiques et de thèses. Il est cependant possible de trouver des données liées à certaines de ces productions scientifiques.

- Les "**data papers**" sont des publications sur les données, souvent en libre accès. Ils augmentent la visibilité et la compréhension de celles-ci avec un descriptif exhaustif. Généralement, ils ne donnent pas accès aux données directement mais indiquent où elles se trouvent via un identifiant pérenne de type DOI.

- Il existe des **moteurs de recherche** spécialisés qui permettent de rechercher des données de recherche comme Google Dataset Search, OpenAlex, DataCite Search...

- Les **entrepôts** sont l'endroit idéal pour trouver des données. Ils sont conçus et dédiés pour le dépôt et le partage des données de recherche. En fonction de vos besoins, il suffit de cibler le bon entrepôt (disciplinaire par exemple) pour trouver les données qui vous intéressent. Pour ce faire, il existe des annuaires tels que re3data et Cat OPIDoR.





4 Parmi ces identifiants, lesquels peuvent s'appliquer aux données ?

Bonnes réponses

- DOI
- Handle
- ARK

Mauvaises réponses

- ORCID
- IdHAL

DOI, Handle et ARK sont tous les trois des identifiants pérennes ouverts (ou « PID » pour Permanent Identifiers) destinés à identifier les données.

ORCID et IdHAL sont des identifiants pérennes pour les auteurs.



- **DOI (Digital Object Identifier)** :
c'est l'identifiant pérenne le plus utilisé pour les données. Au niveau international, l'attribution d'un DOI pour les données de la recherche est gérée par l'organisme DataCite. Pour la France, l'agence d'attribution est l'Inist-CNRS.
- **Handle** : est un autre identifiant pérenne qui référence les objets numériques.
- **ARK (Archival Resource Key)** : identifie tout type d'objets numériques, physiques ou immatériels (concepts).
- **ORCID (Open Researcher or Contributor ID)** : identifiant pour les auteurs et les contributeurs.
- **IdHAL** : identifie un auteur dans l'archive ouverte HAL. Le PID IdHAL ne doit pas être confondu avec "HAL Id" qui identifie la publication dans HAL !



A_{ccessible}



I_{nteroperable}



R_{eusable}



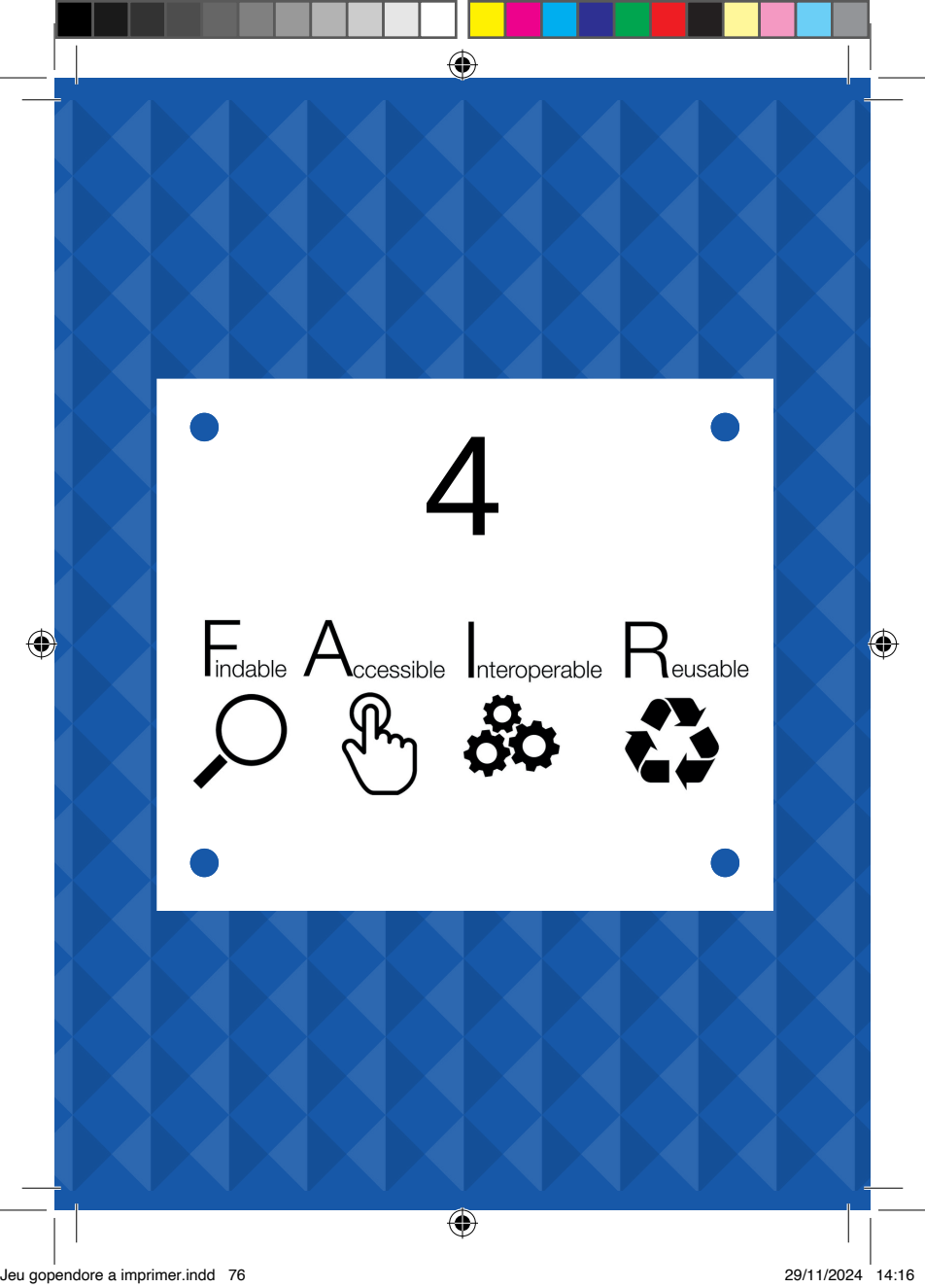
**Parmi ces identifiants,
lesquels peuvent s'appliquer
aux données ?**

(3 bonnes réponses)

- DOI
- ORCID
- Handle
- IdHAL
- ARK

2 points





4

Findable



Accessible

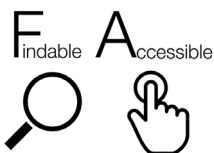


Interoperable



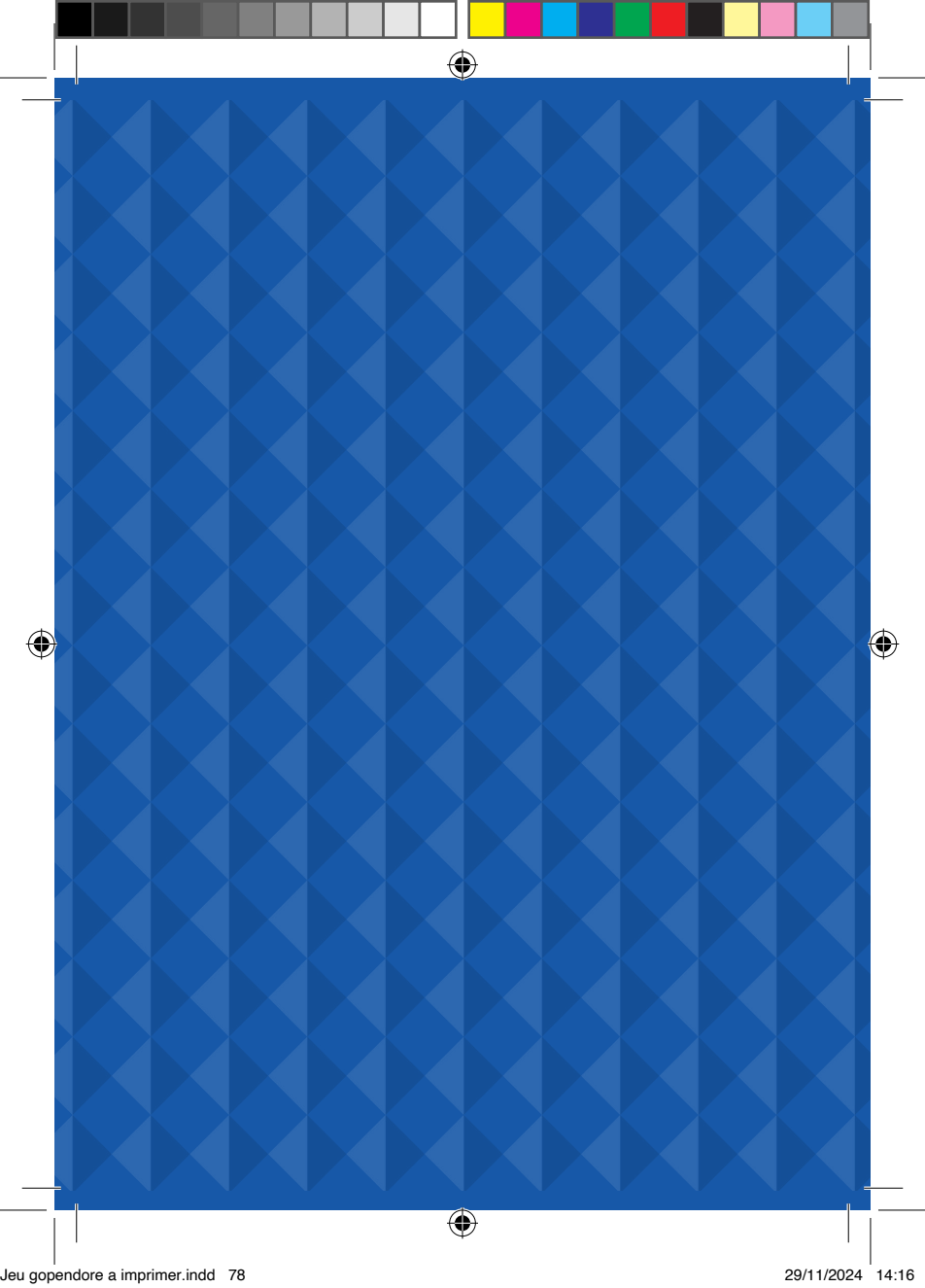
Reusable

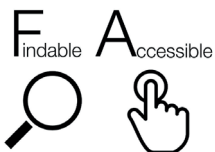




14 L'utilisation de standards et de vocabulaires contrôlés (ontologies, thésaurus) contribue à l'interopérabilité des données.

VRAI ! S'appuyer sur un **lexique** ou un **vocabulaire contrôlé** permet d'optimiser la recherche d'information et l'interopérabilité sémantique entre différents jeux de données : l'échange de données peut ainsi se faire automatiquement d'un système informatique à l'autre.





**L'utilisation de standards
et de vocabulaires contrôlés
(ontologies, thésaurus)
contribue à l'interopérabilité
des données.**

Vrai ou Faux ?

2 points



14

F_{indable}



A_{ccessible}

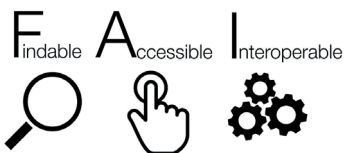


I_{nteroperable}



R_{eusable}





18 Où partager mes données pour permettre leur réutilisation ?

Bonnes réponses

- Dans un entrepôt thématique de confiance
- Dans l'entrepôt national Recherche Data Gouv

Mauvaises réponses

- Sur un réseau social du type Academia ou ResearchGate
- Dans une base de données disciplinaire accessible sur le site web de mon laboratoire



- Il n'y pas d'assurance que le [site web du laboratoire](#) soit connu et pérenne.

- Il n'est pas possible de déposer et donc de partager des données de recherche sur un réseau social, même scientifique, comme [Research Gate](#) ou [Academia](#). Par contre, vous pouvez y signaler l'identifiant pérenne des données déposées par ailleurs dans un entrepôt.

- Il est important de déposer dans un [entrepôt de confiance](#) et de choisir, en fonction de ses données, un [entrepôt thématique ou disciplinaire](#).

- Si aucun entrepôt disciplinaire ou institutionnel ne correspond à vos besoins, vous pouvez déposer vos jeux de données dans l'entrepôt national pluridisciplinaire [Recherche Data Gouv](#).



F



A



I



nteroperable

R



Reusable

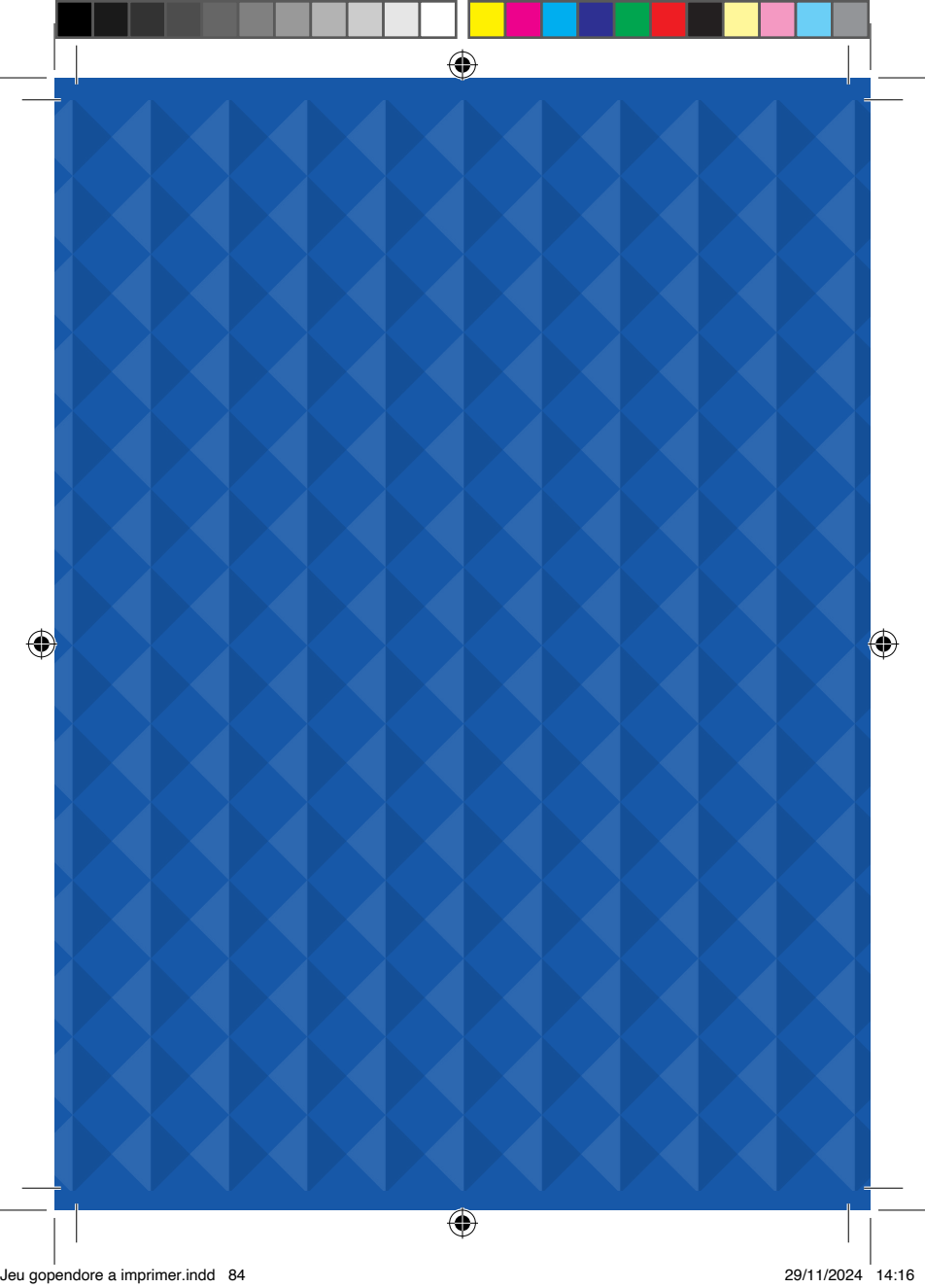
20

Un data paper facilite l'accès aux données.

Vrai ! Un data paper permet l'accessibilité aux données car :

- tous les mécanismes d'accès aux données y sont expliqués précisément
- le lien vers le ou les entrepôts où sont déposées les données est indiqué
- le lien vers le code source du ou des logiciels utilisé(s) pour lire ou traiter les données le cas échéant est également mentionné
- l'identifiant pérenne des données est précisé
- les métadonnées associées y sont décrites.







Accès et visualisation

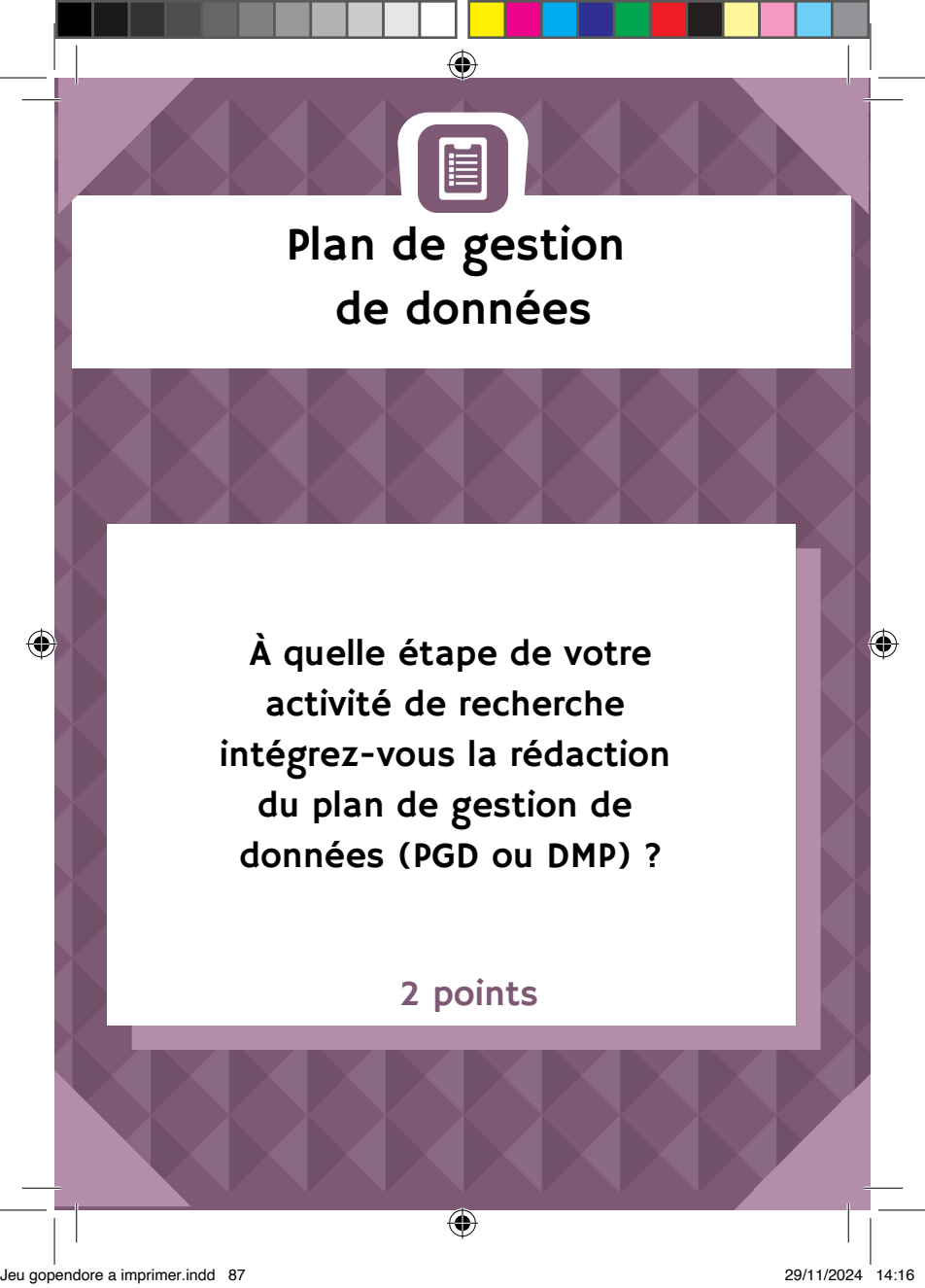
**Que pouvez-vous faire pour valoriser
vos données après le dépôt
dans un entrepôt ?
(3 réponses possibles)**

- les exposer sur un site internet
- les cartographier
- rédiger un data paper
- les archiver définitivement

2 points







Plan de gestion de données

**À quelle étape de votre
activité de recherche
intégrez-vous la rédaction
du plan de gestion de
données (PGD ou DMP) ?**

2 points





Data papers & Data journals

**En quoi un data paper
valorise-t-il vos données ?**

3 points







Aspects juridiques, éthiques, intégrité scientifique

**Quelle(s) licence(s)
est-il recommandé d'attribuer
aux données publiques
en France pour faciliter leur
réutilisation ?**

3 points







Stockage & archivage

**Quelles sont vos pratiques
en ce qui concerne
le nommage des fichiers ?**

I point





Stockage & archivage

**Parmi ces formats de fichiers,
lesquels sont ouverts ?
(3 réponses possibles)**

- .csv
- .pdf
- .doc
- .txt
- .avi

2 points





Dépôt & entrepôts

**Vous déposez vos données
dans un entrepôt.
Comment choisissez-vous
l'entrepôt ?**

2 points





Stockage & archivage

**Comment procédez-vous
pour stocker vos données
durant le projet ?**

I point







Aspects juridiques, éthiques, intégrité scientifique

**Vos données
nécessitent-elles des
autorisations spécifiques
pour leur diffusion ?**

2 points





Dépôt & entrepôts

**Comment préparez-vous
le dépôt de vos données
dans un entrepôt ?**

3 points





Aspects juridiques, éthiques, intégrité scientifique

**Quelle licence Creative Commons
est l'équivalent international
de la licence ouverte française
(Etalab) ?**

- CC-BY**
- CC-BY-NC-SA**
- CCO**

2 points





Identifiants pérennes

**Parmi ces identifiants, lesquels
sont des identifiants auteur ?
(2 réponses possibles)**

- DOI
- ORCID
- IdHAL
- ARK

3 points





Stockage & archivage

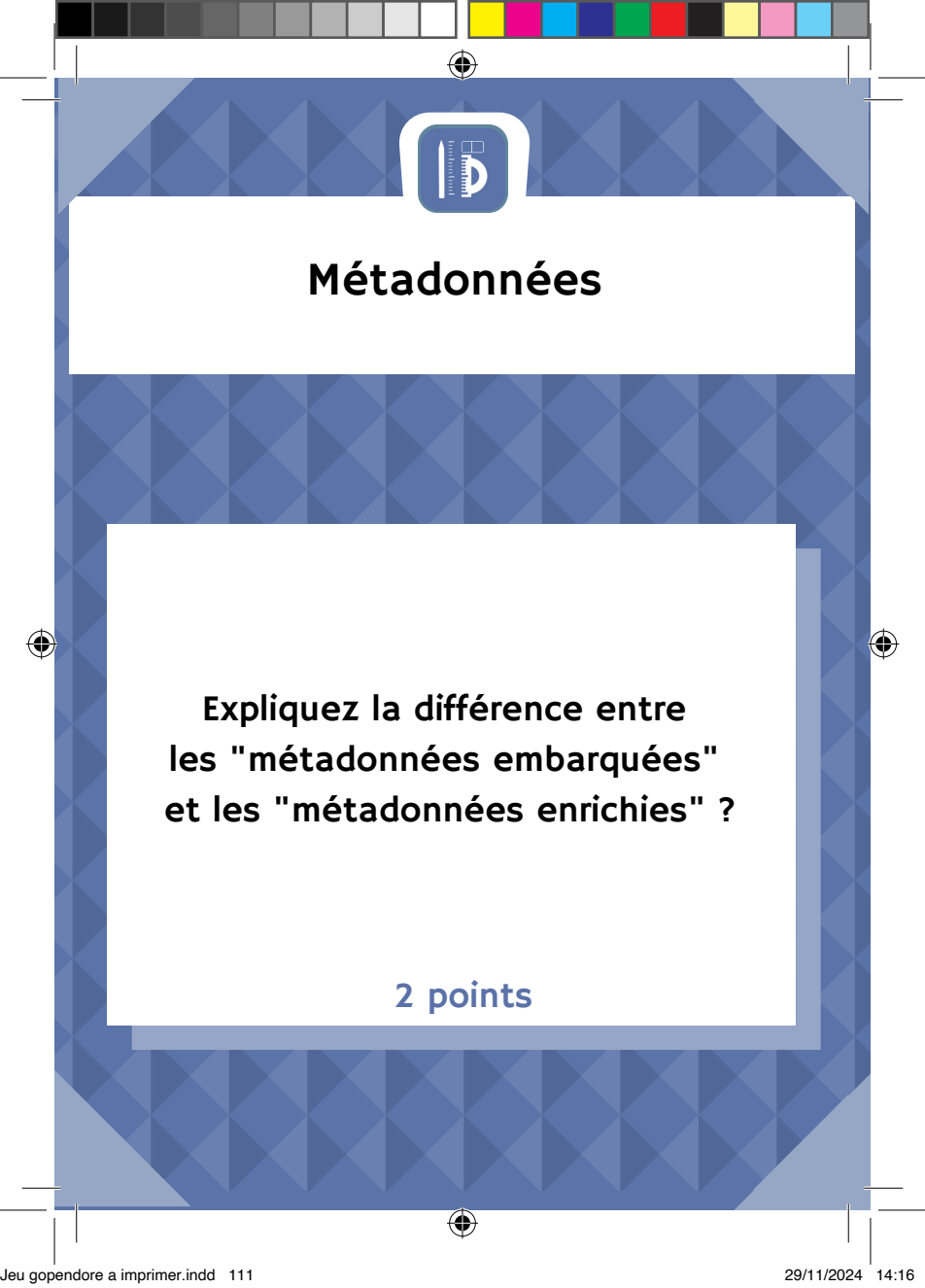
**Si vous utilisez un logiciel spécifique
durant votre projet, que devrez-vous
faire au moment du partage ?**

(2 réponses possibles)

- Garder le format d'origine
- Fournir le nom du logiciel
- Fournir le script

3 points





Métadonnées

**Expliquez la différence entre
les "métadonnées embarquées"
et les "métadonnées enrichies" ?**

2 points





Aspects juridiques, éthiques, intégrité scientifique

**Parmi ces propositions, lesquelles
sont des données sensibles ?**

(2 réponses possibles)

- appartenance ethnique d'une famille dans une étude ethnologique
- géolocalisation d'une espèce végétale protégée
- opinions politiques des personnes dans une interview

2 points





Métadonnées

**Citez un standard de
métadonnées généraliste**

3 points







Métadonnées

**Utilisez-vous un type de
vocabulaire contrôlé spécifique
à votre discipline tel que...**

- Lexique
- Thésaurus
- Nomenclature
- Ontologie

... si oui, lequel ?

2 points







Identifiants pérennes

**Quels sont les avantages à
attribuer un identifiant pérenne
pour vos jeux de données ?**

2 points





Stockage & archivage

**Quelles sont les pratiques
d'archivage
de votre laboratoire ?**

I point







Plan de gestion de données

Que veut dire "DMP" ?

- Digital Management Program
- Debrief en Mode Projet
- Data Management Plan
- Dossier Médical Partagé

I point





Dépôt & entrepôts

**Quel est l'intérêt de
déposer ses données
dans un entrepôt ?**

3 points





Aspects juridiques, éthiques, intégrité scientifique

**Vous avez des données
personnelles à traiter. Savez-vous
à qui faire appel pour vous aider ?**

(2 réponses possibles)

- **Au directeur de l'unité**
- **Au Délégué à la Protection des Données**
- **À un juriste**
- **À un informaticien**
- **À un documentaliste**

3 points





Plan de gestion de données

**Qui peut m'aider à rédiger
un Plan de gestion des données ?
(3 réponses possibles)**

- Un documentaliste**
- Un informaticien**
- Un gestionnaire du laboratoire**
- Le Délégué à la Protection des Données**

I point







Enjeux et bénéfices

**Parmi ces services,
lesquels sont proposés
par l'Inist-CNRS ?**

- Attribution de DOI
- Stockage des données
- Exposition des données
- Aide à la rédaction d'un DMP

2 points





Enjeux et bénéfices

**Quel type d'accompagnement
peut m'apporter un informaticien ?
(3 réponses possibles)**

- Détermination de la propriété des données
- Choix des formats de fichiers
- Sécurité des données
- Stockage des données
- Renseignement des métadonnées

3 points





Stockage & archivage

**Quelles données de recherche
peuvent être éligibles
à l'archivage pérenne ?**

- Toutes les données de recherche
- Les données coûteuses à produire
- Les données dans un format de fichier rare
- Les données non-reproductibles
- Les données à valeur patrimoniale

3 points





Aspects juridiques, éthiques, intégrité scientifique

**Quelle(s) licence(s)
peut-on attribuer
aux logiciels ?**

3 points



27



Aspects juridiques, éthiques, intégrité scientifique

**Quelle(s) licence(s)
peut-on attribuer
aux bases de données ?**

3 points







Métadonnées

**Citez un standard de
métadonnées spécifique
à votre discipline**

3 points





29



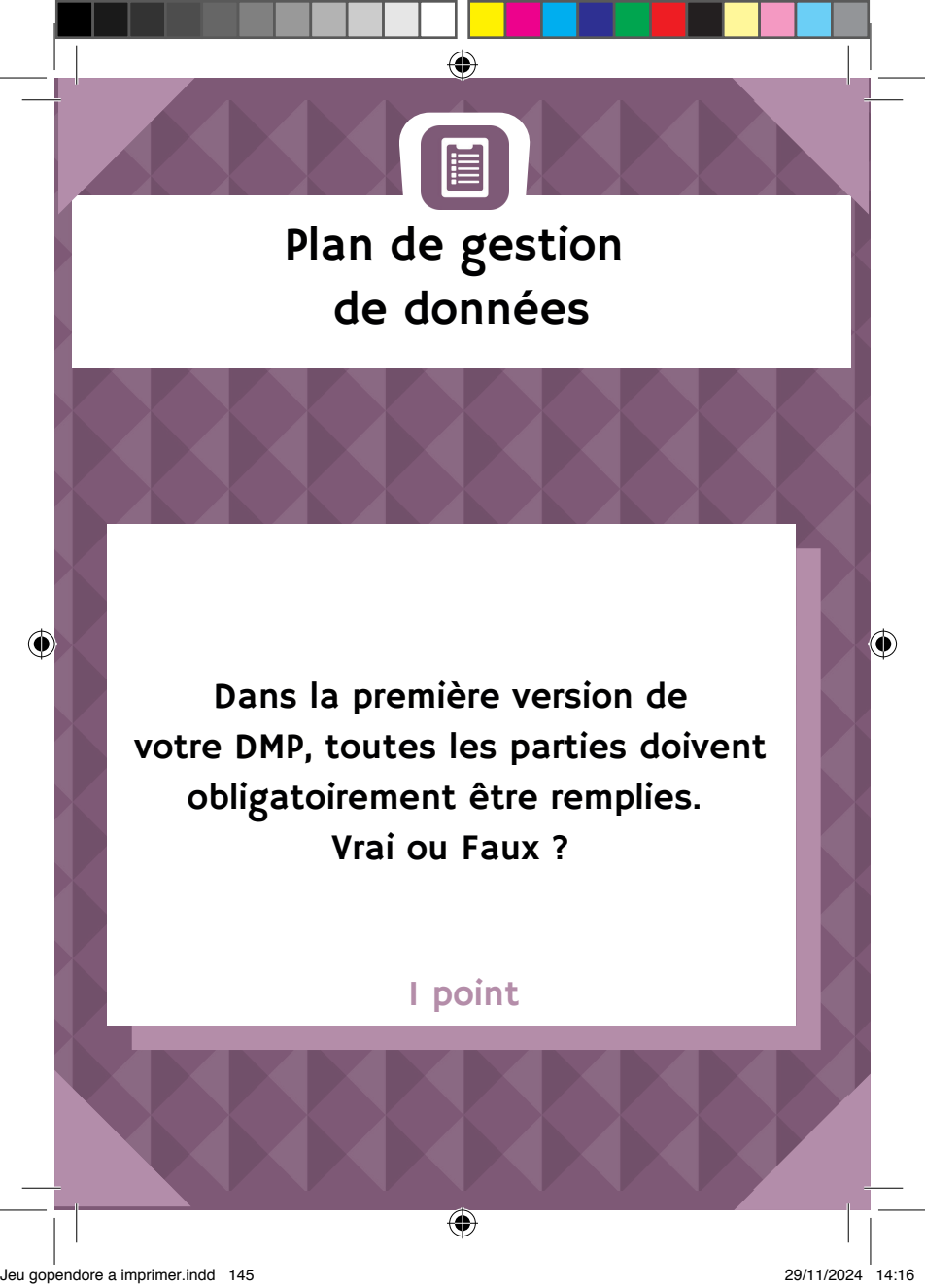
Plan de gestion de données

Que veut dire "PGD" ?

- Prévisionnel de Gestion de Données
- Plan de Gestion des Données
- Programme de Gestion des Données

I point





Plan de gestion de données

**Dans la première version de
votre DMP, toutes les parties doivent
obligatoirement être remplies.**

Vrai ou Faux ?

1 point





Dépôt & entrepôts

**Citez un
entrepôt de données
généraliste**

2 points





Dépôt & entrepôts

**Citez un
entrepôt de données
disciplinaire**

3 points

