

## Atelier sur les données de recherche

Plan de Gestion de Données (PGD) : retour d'expérience projets de laboratoire

Jean-Philippe TALEC

CNRS IKER (UMR 5478), Documentation

13 avril 2017, Pau



# Sommaire du document

- 1 Avant-propos
- 2 Introduction générale
- 3 Structure : sections et éléments
- 4 Exemples
- 5 Bonnes pratiques
- 6 Exercices
- 7 Bibliographie
- 8 Annexe

# Sommaire de section

- 1 Avant-propos
  - Animateur de l'atelier
  - Centre de recherche sur la langue et les textes basques IKER
  - Objectifs pédagogiques de l'atelier
- 2 Introduction générale
- 3 Structure : sections et éléments
- 4 Exemples
- 5 Bonnes pratiques
- 6 Exercices
- 7 Bibliographie
- 8 Annexe

# Animateur de l'atelier

## Fonctions actuelles

- Ingénieur d'études (IE) CNRS (2002-....) au centre de recherche sur la langue et les textes basques IKER (2007-.....)
  - Responsable du centre de documentation de l'unité
  - Administrateur du Système d'information documentaire (SIGB) KOHA
  - Impliqué dans des projets de recherche (Contrat quinquennal 2016-2020 Maison des Sciences de l'Homme d'Aquitaine (MSHA))
  - Responsable de la version électronique de la revue d'études basques *Lapurdum*
  - Chargé de cours Licence-Master au Département inter-universitaire d'Études Basques de Bayonne (DIEB)

## Coordonnées

- Courriel : [jean-philippe.talec@iker.cnrs.fr](mailto:jean-philippe.talec@iker.cnrs.fr), Tél. : 05.59.57.42.02 (bureau)

# Centre de recherche sur la langue et les textes basques IKER

Le laboratoire IKER UMR 5478, situé à Bayonne, constitue la seule structure en France spécialisée dans le domaine des études basques.

## Ancrage institutionnel de l'UMR (1999-....)

- ① le Centre National de la Recherche Scientifique (CNRS)
- ② l'Université de Bordeaux-Montaigne (UBM)
- ③ l'Université de Pau et des Pays de l'Adour (UPPA)

## Équipe de recherche

- ① 7 Chercheurs CNRS spécialistes du basque, de l'occitan et du breton
- ② 4 Enseignants-chercheurs (UBM, UPPA)
- ③ 4 ITA (1 en BAP F, 1 en BAP D, 2 en BAP J)
- ④ 21 Doctorants dont la moitié réalise des thèses en sciences du langage

## Source

- CNRS IKER (UMR 5478). *Site Web du laboratoire*. Disponible à l'adresse : <<http://www.iker.cnrs.fr>> [consulté le 17.02.2017]
- Documentation IKER. *Présentation des ressources et services documentaires*. Disponible à l'adresse : <<http://www.iker.cnrs.fr/-centre-de-documentation-.html>> [consulté le 17.02.2017]

# Collections documentaires et services de soutien aux chercheurs

Créé en 2009, le centre de documentation vise à soutenir les personnels de l'unité, les masters d'études basques et un certain nombre de visiteurs (chercheurs invités).

## Équipements/Collections

- salle de lecture de 100 m<sup>2</sup>, wifi, postes informatiques en libre accès
- 6.000 documents imprimés en libre accès (centaine d'achats/an)
- importante documentation numérique (*ebooks*, données de recherche) sur des serveurs dédiés et sécurisés

## Services documentaires/numériques

- accompagnement dans la préparation/suivi de projets de recherche
  - rédaction des parties spécifiques au stockage de données, valorisation de résultat de recherche
  - acquisition et mise à disposition de ressources (livres, *ebooks*)
  - recherches et fourniture de documents (article) sur les plateformes documentaires
  - création de sites web (bibliothèque numérique ANPERSANA)

## Sources

- Documentation IKER. *Données de recherche ANPERSANA*. Disponible à l'adresse : <https://anpersana.iker.univ-pau.fr> [consulté le 17.02.2017]
- ——. *Catalogue documentaire KUTXA*. Disponible à l'adresse : <http://iker-koha.univ-pau.fr> [consulté le 17.02.2017]
- ——. *Revue Lapurdum sur Revues.org*. Disponible à l'adresse : <https://lapurdum.revues.org> [consulté le 17.02.2017]

# Objectifs pédagogiques de l'atelier Plan de Gestion de Données (PGD)

Au terme de l'atelier, il serait souhaitable que le participant soit en mesure de :

- comprendre la démarche de création d'un PGD
- décrire et caractériser les parties d'un PGD (sections, éléments, informations)
- partager ses expériences, ses pratiques de gestion de données numériques
- identifier un certain nombre de plateformes de dépôt de données de la recherche
- passer de la conception à la réalisation d'un PGD

## Étude de cas et exercice

- PGD La correspondance en langue basque du *Dauphin* (1757) dans le contexte du Contrat quinquennal 2016-2020 de la Maison des Sciences de l'Homme d'Aquitaine (MSHA) (Programme thématique : 1. Territoires 3. Europe et espace monde 5. Humanités numériques)
- 4 Exercices : contenus du PGD, (bonnes) pratiques, plateformes de données de la recherche, outils en ligne

# Couverture du PGD *le dauphin* et bibliothèque numérique ANPERSANA



Versión 1.1.14, Janvier 2017

**Plan de Gestion de Projet (PGD)**  
collection immatérielle *Le DAUPHIN*

<b>Organisation :</b>	CHRS IBER (UMR 5476)
<b>Facteur :</b>	christophe@scd.uppa.fr
<b>Référence :</b>	scd/04/04/01/dauphin/pgd/
<b>Participants :</b>	équipe de travail programme Le Dauphin
<b>Diffusion :</b>	Participants au groupe de travail Le Dauphin
<b>Date réalisation :</b>	Février 2015
<b>Etat :</b>	Valide

CHRS IBER (UMR 5476)  
Département UFR 11, 16000 Mont de Marsan, Bayonne  
T. 053009 58 57 42 02  
<http://www.iber.univ-bordeaux-montp.fr/centre-de-documentation-iber/>



Bibliothèque numérique  
**Anpersana**



- Parcourir les contenus
- Parcourir les collections
- Sous le cerner
- Carte
- Nous contacter
- A propos

**Focus sur ...**

Quelques Extraits

**VOCABULAIRE BASQUE**

par M<sup>r</sup> GUILBEAU

[Sur l'enseignement des langues régionales](#)

Cette étude, avec quelques différences de forme et divisée en quatre articles, a déjà paru dans le journal *Le Courrier de Bayonne* (n° 23 et 30...

[Collection mise en avant](#)

**Contenus ajoutés récemment**

**1908CAREN** urtheho estuakidunentzat estuzakoa eguzkiti edo Almanaka Berria

L'almanach est publié une fois par an. Il reprend les grandes dates du calendrier, les fêtes religieuses et éponymes. Le numéro de 1908...

[Quelques extraits du vocabulaire basque régionalisé](#)

Martin Guilbeau (1839-1912) traduit des termes de la langue basque (dialecte labourdin) du début du XIXe siècle en français.

[Sur l'enseignement des langues régionales](#)

Cette étude, avec quelques différences de forme et divisée en quatre articles, a déjà paru dans le journal *Le Courrier de Bayonne* (n° 23 et 30...



# Sommaire de section

- 1 Avant-propos
- 2 **Introduction générale**
  - Contexte européen
  - Définition des concepts de "données de la recherche" et de "PGD"
  - Bénéfices pour le chercheur ou le chargé de projet de recherche
- 3 Structure : sections et éléments
- 4 Exemples
- 5 Bonnes pratiques
- 6 Exercices
- 7 Bibliographie
- 8 Annexe

# Ouverture des données de la recherche (programme Horizon H2020)

De plus en plus la Commission Européenne (CE) conditionne le financement de programmes de recherche au partage de données collectées ou produites sur les fonds publics (mouvement de l'*Open Access* et de l'*Open Data*).

## Caractéristique

- L'Enseignement Supérieur et la Recherche (ESR) Français considère les données comme un gisement potentiel d'innovations permettant la croissance économique.
- Il s'engage dans l'ouverture et la réutilisation des données :
  - la Bibliothèque scientifique numérique (BSN) a un segment, BSN 10 "Données de la recherche".
  - La Direction de l'information scientifique et technique (DIST) du CNRS inclut la gestion des données de la recherche dans ses axes stratégiques de l'Information scientifique et technique (IST).

## Sources

- Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche (MESR). *Programme H2020*. Disponible à l'adresse : <<http://www.horizon2020.gouv.fr/cid82025/1e-libre-acces-aux-publications-et-aux-donnees-de-recherche.html>> [consulté le 10.02.2017]
- Inist-CNRS. *BSN 10 : données de la recherche*. Disponible à l'adresse : <<http://www.bibliothequescientifiquenumerique.fr/bsn-10-donnees-de-la-recherche>> [consulté le 10.02.2017]
- DIST du CNRS. *Le travail de la science et le numérique : données, publications, plateformes*. Disponible à l'adresse : <[http://www.cnrs.fr/dist/z-outils/documents/20170203\\_analyse%20syst%C3%A9mique\\_vf.pdf](http://www.cnrs.fr/dist/z-outils/documents/20170203_analyse%20syst%C3%A9mique_vf.pdf)> [consulté le 8.02.2017]
- PARKER T. H., NAKAGAWA S. GUREVTICH J.. Open data : towards full transparency. *Nature* 538, 27 octobre 2016. DOI : <10.1038/538459d> [consulté le 22.02.2017]

# Données de la recherche

## Caractéristiques

- Les données de recherche renvoient le plus souvent de faits ou de nombres constitués en vue d'être examinées et considérées comme base de raisonnement à des discussions scientifiques.
- Il existe une définition plus précise formulée par l'OCDE : "*Enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche*"
- Les données de la recherche recouvrent un périmètre large :
  - données primaires ou brutes : information enregistrée (textes, images, données d'observation)
  - données issues d'un traitement particulier : analyses, descriptions, transcriptions...
- Les jeux de données peuvent devenir de nouveaux instruments de la science.

## Source, compléments

- OCDE. *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics* (2007). Disponible à l'adresse : <http://www.oecd.org/fr/sti/sci-tech/38500823.pdf> [consulté le 10.02.2017]

# Plan de Gestion de Données (PGD), *Data Management Plan (DMP)*, *Data Plan Management (DPM)*

## Caractéristiques

- C'est un document formalisé décrivant la manière dont seront collectées, traitées et (re)utilisées les données produites au cours et à l'issue du processus de recherche.
- L'objectif est de regrouper les nombreux aspects régissant le cycle de vie des données de la recherche dans un document unique, auparavant dispersés entre divers acteurs et documents (carnet de route).
- Le PGD doit préciser les choix réalisés en termes de :
  - description des données collectées et/ou créées,
  - prévisions de sécurité d'accès et partage de données,
  - stratégie de la préservation à long-terme (archivage),
  - questions d'ordre éthiques, de propriété intellectuelle et de restrictions,
  - coûts associés à la gestion et conservation des données.
- Le PGD est un livrable pour les participants au groupe de recherche (intervenants sur le projet).

# Aspects à développer dans un PGD

## Questions à poser pendant la création d'un PGD

- Quelles données seront créées (format, types, volume ...) ?
- Quelles normes et méthodologies seront à utiliser (y compris les métadonnées) ?
- Comment la propriété intellectuelle des données sera traitée ?
- Quelle sera la stratégie de partage et d'accès aux données, la préservation à long-terme (archivage) ?

## Point de vigilance

- Le PGD doit être mis à jour tout au long du projet de recherche.
- En France, il n'y a pas pour le moment d'obligation des financeurs de rédiger un PGD dans une réponse à un appel à projet national (type ANR) ou régional (Contrat État-Région).
- En principe l'institution demandant un PGD fournit une trame dont l'utilisation est fortement recommandée (projet européen H2020).
- La description des données varient selon les disciplines (biologie, sciences de la terre, physique, shs...).

# Organisations commanditaires de PGD

Les organismes de financement ont leur politique de gestion des données propre (différences entre PGD).

## Liste loin d'être exhaustive

- Amérique du nord : National Science Foundation (NSF), National Aeronautics and Space Administration (NASA), US Department of Energy, National Research Council Canada...
- Europe : Research Councils UK, University of Glasgow, Medical Netcare GmbH, Commission européenne...

## Source

- European Commission. *Data management plan - general definition*. Disponible à l'adresse : <[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)> [consulté le 17.02.2017]

# Bénéfices pour le chercheur ou le chargé de projet de recherche

## Outil pédagogique

- 45-60 questions structurées et détaillées sur la gestion des données de la recherche aidant à la construction du projet (début du projet)
- les réponses du PGD vont éventuellement consolider les questionnements scientifiques (prise en considération des données dès le début du projet)

## Outil de recherche

- mise à disposition des données pour reproduire les expériences (ré-utilisation)
- utilisation libre de données produites par d'autres laboratoires
- amélioration de la qualité des données (traçabilité des actions sur les données) et des conditions d'archivage





# Sommaire de section

- 1 Avant-propos
- 2 Introduction générale
- 3 Structure : sections et éléments**
  - Sections du document
  - Éléments du document
- 4 Exemples
- 5 Bonnes pratiques
- 6 Exercices
- 7 Bibliographie
- 8 Annexe

# Sections

Un PGD est composé de plusieurs sections.

## Structure

- Section 1 : informations sur le projet
- Section 2 : responsabilités des données
- Section 3 : ressources nécessaires à la mise en oeuvre du PGD
- Section 4 : description des données
  - Section 4.1 : description du jeu de données
  - Section 4.2 : stockage, accès, sécurité au cours du projet
  - Section 4.3 : métadonnées (documentation et organisation des données)
  - Section 4.4 : à l'issue du projet diffusion, partage et réutilisation des données
- Section 5 : sélection et archivage

# Éléments I

## Liste d'éléments recommandés

Éléments	Description	Section
Description des données	identifiant et nom du jeu de données ; type (enquête, corpus de texte, photographies, code logiciel...) et méthodes de production des données	4.1
Données existantes	présentation des données existantes éventuellement produits par d'autres équipes de recherche ; procédures d'intégration des données au projet	4.2
Format	indications sur les formats dans lesquels les données seront produites, converties, stockées et mises à disposition	4.1
Métadonnées	description des métadonnées à fournir avec les données générées, citation de la norme ou standard (de métadonnées) à utiliser	4.3

# Éléments II

## Liste d'éléments recommandés

Éléments	Description	Section
Stockage et sauvegarde	procédures de stockage et de sauvegarde des données ; description des volumes, supports, type d'hébergement des données de recherche	4.2
sécurité	description des risques susceptibles de menacer les exigences de confidentialité, de disponibilité et d'intégrité des données tout au long de la phase active du projet (autorisations, restrictions, embargos)	4.2
Responsabilité	noms des personnes responsables de la gestion des données dans le projet de recherche	2
Propriété intellectuelle	noms des organismes ou personnes ayant les droits de propriété sur les données et la manière dont la propriété intellectuelle sera protégée si nécessaire	2

# Éléments III

## Liste d'éléments recommandés

Éléments	Description	Section
Accès et partage	description de la façon dont les données seront partagées, des procédures d'accès ouverts ou accordés uniquement à des groupes d'utilisateurs spécifiques ; fourniture d'un échéancier pour le partage de données et la publication de la ressource	4.2
Public	description des utilisateurs potentiels secondaires des données	4.4
Durée de conservation	description de la période et de la façon dont les données seront conservées	5
Archivage et préservation	procédures en place ou envisagées pour l'archivage à long terme et la préservation des données	5
Éthique et vie privée	description de la façon dont sera traité et protégé la vie privée (déclaration à la CNIL, consentement des informateurs, anonymisation des données)	4.4

## Éléments IV

## Liste d'éléments optionnels

Éléments	Description	Section
Budget	description des coûts de préparation des données, des espaces de stockage et de la plateforme d'archivage	3
Exigences légales	liste des exigences pour la gestion des données et le partage des données	5

# Sommaire de section

- 1 Avant-propos
- 2 Introduction générale
- 3 Structure : sections et éléments
- 4 Exemples**
- 5 Bonnes pratiques
- 6 Exercices
- 7 Bibliographie
- 8 Annexe

Section 1 : PGD *Le Dauphin*

1. Informations sur le projet / 1- Informations on the project	
<p>Cette section a pour vocation de renseigner administrativement sur le projet auquel ce PGD est lié, de le présenter et de le décrire succinctement. [This section aims to provide administrative information about the project to which this DMP is linked and to present and briefly describe it.]</p>	
Référence de la convention de financement, Grant agreement number	Contrat quinquennal 2016-2020 Maison des Sciences de l'Homme d'Aquitaine (MSHA)
Programme de recherche, Research program	Programme thématique: 1. Territoires 3. Europe et espace monde 5. Humanités numériques
Acronyme du projet, Project acronym	Le DAUPHIN
Titre du projet, Project title	La correspondance en langue basque du <i>Dauphin</i> , 1757
Objectifs du projet, Goals of the project	<p>La démarche générale de ce projet est de collecter ces lettres actuellement dispersées à Londres, d'en faire l'annotation, d'en fournir une édition pour les mettre à disposition des chercheurs et du public sur une plateforme informatique qui en assurera l'archivage pérenne. Ce corpus sera matière à traitement multidisciplinaire du point de vue linguistique et historique. Le projet est organisé selon trois objectifs:</p> <ol style="list-style-type: none"> <li>1. Confection et mise à disposition d'un corpus annoté</li> <li>2. Exploitation scientifique du point de vue historique (histoire des transports, des idées, nouvelle histoire atlantique)</li> <li>3. Exploitation scientifique du point de vue linguistique (exploitations philologique, sociolinguistique, littéraire)</li> </ol>
Mots-clés du projet, Keywords	bateau; commerce, communications, transports; correspondance; mer; famille; guerre de sept ans; basque (langue); histoire du Pays basque; transport maritime; linguistique
Coordinateur/Bénéficiaire, Coordinator/Recipient	CNRS IKER (UMR 5478)
Responsable scientifique du projet, Project leader	Charles VIDEGAIN <a href="mailto:charles.videgain@univ-pau.fr">charles.videgain@univ-pau.fr</a>
Affiliation et unité de rattachement du responsable scientifique, Administrative affiliation of the project leader	IKER UMR 5478 CNRS, Université UPPA
Document référence	Réponse à l'appel à projet : Videgain_LeDauphin.pdf



Section 2 : PGD *Le Dauphin*

## 2. Responsabilité des données / 2. Responsibility for the data

Cette section vise à identifier la ou les personne(s) qui seront en charge de la mise en œuvre et de la mise à jour du PGD tout au long du projet, ainsi que la propriété intellectuelle des données liées au projet de recherche. / *This section provides information about the person(s) who will be responsible for implementing and updating the DMP throughout the project, and the intellectual property of them.*

**Nom du responsable de la gestion des données au cours du projet de recherche / Name of the person in charge of data management during the project**

Jean-Philippe TALEC  
[jean-philippe.talec@iker.cnrs.fr](mailto:jean-philippe.talec@iker.cnrs.fr)

**Propriété des données / Data property**

A la fin du projet, le laboratoire partagera la plupart des données de recherche qu'il produira (lettres photographiées, transcriptions en basque et métadonnées Dublin Core), à tous sur un site web, en format ouvert, avec liberté de réutilisation sous licence *Creative Commons* (CC- BY-NC-SA).

- **BY (ATTRIBUTION)** : Toutes les licences CC obligent ceux qui utilisent vos œuvres à vous créditer de la manière dont vous le demandez, sans pour autant suggérer que vous approuvez leur utilisation ou leur donner votre aval ou votre soutien.
- **NC (PAS D'UTILISATION COMMERCIALE)** : Vous autorisez les autres à reproduire, à diffuser et à modifier votre œuvre, pour toute utilisation autre que commerciale, à moins qu'ils obtiennent votre autorisation au préalable.
- **SA (PARTAGE DANS LES MEMES CONDITIONS)** : Vous autorisez les autres à reproduire, diffuser et modifier votre œuvre, à condition qu'ils publient toute adaptation de votre œuvre sous les mêmes conditions que votre œuvre. Toute personne qui souhaiterait publier une adaptation sous d'autres conditions doit obtenir votre autorisation préalable.

La base de données ANPERSANA est la propriété du laboratoire.

## Section 4 : PGD *Le Dauphin*

<b>Nombre de jeu(x) de données/ Number of datasets</b>	Un jeu de données sera produit dans le cadre du projet <i>Le Dauphin</i> . Il correspondra à des documents collectés aux archives de l'Amirauté britannique ( <i>High Court Admiralty</i> ) et générés pendant la durée du projet.
--	--

### 4.1. Description des données – Jeu de données

#### 4.1. Data description – Dataset [en] /

Cette section a pour vocation de présenter le jeu de données qui sera produit et/ou reçu dans le cadre du projet.

*This section aims to generally present the type of data that will be produced and/or received in connection with the project.*

<b>Identifiant et nom du jeu de données / Reference and name of data set</b>	Dataset 2016 ledauphin \\nas-iker\ ledauphin
<b>Nature des données / Nature of data</b>	Les données produites seront iconographiques et textuelles avec un corpus de photographies et de transcriptions. Les textes présenteront un lexique basque utilisé au Pays basque au 18 <sup>e</sup> siècle.
<b>Réutilisation de données existantes/ Reuse of existing data</b>	Les originaux des lettres sont conservés dans les archives de l'Amirauté britannique ( <i>High Court Admiralty</i> ) (HCA), 30/264/2) à Londres. Cette structure a autorisé l'équipe de recherche à publier les images numériques des lettres.  Licence CC-BY-SA
<b>Méthode de production des données/ Method of production of data</b>	La production des lettres numérisées se fera avec un appareil photographique, les transcriptions avec un logiciel de traitement de texte et les annotations avec un éditeur XML. Ce jeu de données sera intégré dans la base de données ANPERSANA fonctionnant avec le logiciel OMEKA version 2.4 et accessible à travers une interface web.
<b>Formats des données / Data standard</b>	Les formats des données seront ouverts afin de faciliter le partage.  Photographies originales : JPG



# Sommaire de section

- 1 Avant-propos
- 2 Introduction générale
- 3 Structure : sections et éléments
- 4 Exemples
- 5 Bonnes pratiques**
  - Acteurs, étapes, planning
  - Guide de stockage, sauvegarde, archivage, documentation
- 6 Exercices
- 7 Bibliographie
- 8 Annexe

# Intervenants/compétences dans la rédaction du PGD

La rédaction du PGD revient à l'équipe de recherche, aidée d'acteurs susceptibles d'accompagner la mise en oeuvre du projet.

## Coordinateur, responsable principal de la rédaction

- Chercheur, informaticien, professionnel de l'IST
  - création et rédaction de la première version
  - intégration des corrections/modifications du groupe de travail
  - rédaction de la version finale

## Contributeur du PGD

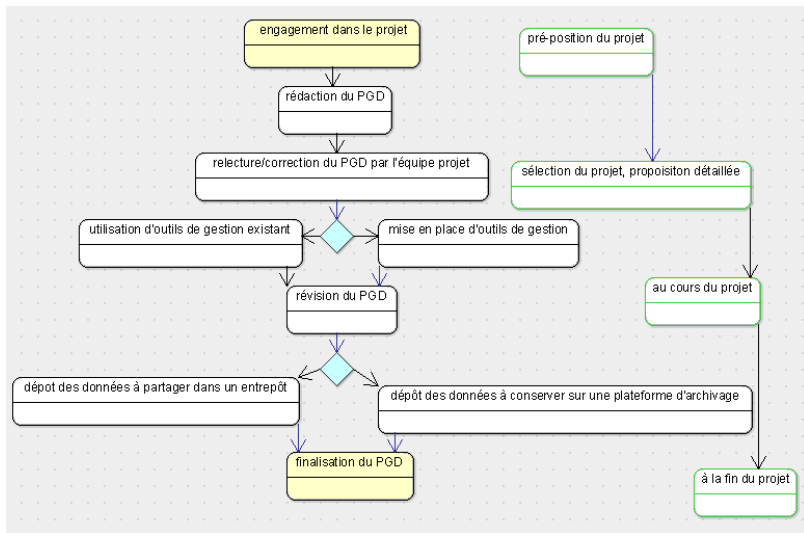
- Chercheur : producteur et utilisateur de la recherche
  - expertise sur le contenu informationnel et les usages des données
- Ingénieur projet : montage du projet
  - expertise sur la conduite du projet (délai, budget, aspect organisationnel)
  - service d'appui à la recherche et à l'innovation dans les universités, pôle Europe et International du CNRS
- Informaticien : sécurité des données
  - expertise sur les systèmes d'information et les les outils à utiliser
  - pôle numérique de l'université, direction des systèmes d'information (DSI)
- Professionnel de l'IST : cycle de vie des données
  - expertise sur l'organisation des données et les métadonnées associées
  - service de documentation, archives, bibliothèque
- Juriste
  - expertise sur la propriété et les licences d'utilisation des données

# Aperçu des étapes de travail

## 7 étapes

- 1 choisir un modèle de licence pour le partage et la ré-utilisation de données (nécessite de clarifier la question de la propriété des données)
- 2 constituer le jeu de données (nécessite de disposer d'équipements informatiques)
- 3 modéliser le PGD ou utiliser celui de l'organisme de financement
- 4 décrire l'organisation des dossiers et fichiers
- 5 documenter les sections du PGD (à partir de métadonnées, de guides de bonnes pratiques, référentiels)
- 6 organiser le partage des données et la sélection des données à publier
- 7 publier le jeu de données

# Planning (liste des tâches)



# Illustration

```

(())var c=function(b){this.element=a(b)};
wn-menu")\d=b.data("target");if(d||(d=b.attr("href"),d=d&&d.replace(/.*(?:#[\s\S]+)?/))
),f=a.Event("hide.bs.tab",{relatedTarget:b[0]}),g=a.Event("show.bs.tab",{relatedTarget:
Prevented(){var h=a(d);this.activate(b.closest("li"),c),this.activate(h,h.parent(),fun
er({type:"shown.bs.tab",relatedTarget:e[0]}))}}},c.prototype.activate=function(b,d,e){
active").removeClass("active").end().find('[data-toggle="tab"]').attr("aria-expanded",
xpanded",!0),h?(b[0].offsetWidth,b.addClass("in"):b.removeClass("fade"),b.parent(".dro
nd('[data-toggle="tab"]').attr("aria-expanded",!0),e&&e())var g=d.find("> .active"),h=
||!d.find("> .fade").length);g.length&&h?g.one("bsTransitionEnd",f).emulateTransitionE
d=a.fn.tab;a.fn.tab=b,a.fn.tab.Constructor=c,a.fn.tab.noConflict=function(){return a.
");(document).on("click.bs.tab.data-api",[data-toggle="tab"],e).on("click.bs.tab.
trict";function b(b){return this.each(function(){var d=a(this),e=d.data("bs.affix"),f
of b&&e[b]()}}var c=function(b,d){this.options=a.extend({},c.DEFAULTS,d),this.$target
proxy(this.checkPosition,this)).on("click.bs.affix.data-api",a.proxy(this.checkPositi
this.pinnedOffset=null,this.checkPosition());c.VERSION="3.3.7",c.RESET="affix affix-
e=function(a,b,c,d){var e=this.$target.scrollTop(),f=this.$element.offset(),g=this.$
tom"--this.affixed)return null!=c?(e+this.unpin<=f.top)&&"bottom":!(e+g<=a-d)&&"bot
&&c?"top":null!=d&&i+j>a-d&&"bottom"},c.prototype.getPinnedOffset=function(){if(
ET).addClass("affix");var a=this.$target.scrollTop(),b=this.$o
EventLoop=function(){setTimeout(a.proxy(th
eight(),d=this.optio

```

: <https://pixabay.com/fr/users/3844328-3844328> :



# Bonnes pratiques stockage, sauvegarde

## Stockage

- Les ordinateurs et les serveurs partagés peuvent être de bons endroits pour le stockage temporaire de vos fichiers fonctionnels.
- Pour un stockage à long terme, les données doivent être mises en place dans des systèmes de préservation bien gérés (DN UPPA, équipement laboratoire).
- Les disques externes sont à utiliser uniquement pour le transfert de fichiers (trop faciles à perdre).
- Il est important d'enregistrer des copies de données dans des formats ouverts et stables (par exemple, ascii, .txt, .csv, .pdf) pour une accessibilité à long terme.
- Il est important de garder une copie du format d'origine, en raison des pertes potentielles de formatage pendant la conversion.

## Sauvegarde

- Règle de 3 : conserver 2 copies sur site, 1 hors site.
- Sauvegarde régulièrement et fréquente - automatiser le processus si possible.

# Bonnes pratiques archivage

## Archivage

- Préférer le stockage centralisé (consulter les informaticiens).
- Archiver les données brutes et les produits intermédiaires / dérivés qui sont coûteux à reproduire ou peuvent être directement utilisés pour l'analyse.
- Archiver tout code script utilisé pour nettoyer et transformer les données brutes.
- Archiver le logiciel utilisé pour le travail sur les données.
- Enregistrer les données de tableaux dans un format de texte délimité.
- Enregistrer les données dans les formats non compressés et non cryptés, si possible.

## Sauvegarde

- CINES. *Archivage numérique pérenne*. Disponible à l'adresse :   
 <<https://www.cines.fr/archivage>> [consulté le 6.04.2017]

# Bonnes pratiques organisation

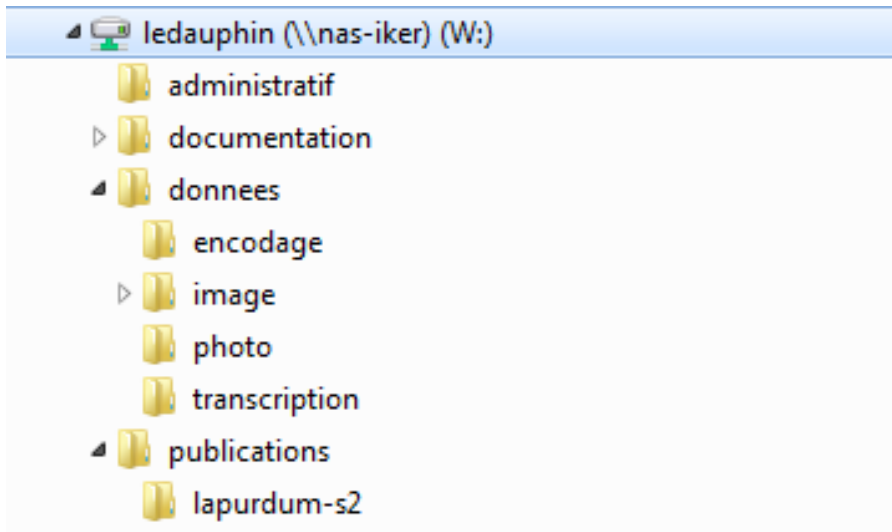
## Organisation de fichiers et de dossiers

- Choisir un système de classement cohérent qui aurait encore du sens pour vous ou pour quelqu'un d'autre dans cinq ans.
- Choisir une hiérarchie logique de répertoires, par exemple, "`<acronyproject-transcription-date>`", "`<enquêteur-processus-date>`" ou "`<instrument-date-echantillon>`".

## Dénomination de fichiers

- Affecter les noms de fichiers descriptifs avec des aspects significatifs de l'étude (quoi, quand, où, comment, qui, pourquoi, version).
- "`<interviewiker – jon20170413`" > " `<cell – helatraitement05,ep001`" >
- Utiliser des conventions de nommage particulières si la discipline le recommande.
- utiliser des lettres majuscules ou des soulignements entre mots, plutôt que des espaces.
- utiliser le format de la date "`<AAAA-MM-JJ>`", ou "`<AAAAMMJJ>`".
- utiliser le numéro de version "`<fichierv001`">
- Utiliser des outils gratuits de renommage de fichiers disponibles : Bulk Rename Utility, Renamer, PSRenamer

# Illustration : organisation des données *Le Dauphin*



# Bonnes pratiques documentation

## Document de travail du projet de recherche

- Au démarrage du projet, créer un fichier (texte ou une feuille de calcul) avec des informations sur les membres de l'équipe de recherche, leurs fonctions, les sources de financement, les méthodes, techniques et normes utilisées, les matériel et logiciels (versions), les sources collectées.
- Y ajouter toutes les restrictions appliquées à la diffusion ou à l'utilisation des données (informations de base sur le projet que l'on pourrait oublier après la fin du projet).
- Maintenir une liste des documents associés au jeu de données (réponse appel à projet, budget).

## Fichiers de données du projet de recherche

- Prendre des notes sur les changements de fichiers, les changements de nom, les dates des modifications.
- Conserver un enregistrement des modifications sous la forme d'un texte (README.txt) ou d'une feuille de travail dans un fichier externe situé dans le même répertoire que les données.
- Conserver le fichier de données non corrigé avec toutes ses erreurs.
- Utiliser des scripts pour les corrections des fichiers (nettoyage). Commenter le script, afin que de mémoriser les analyses effectuées.
- Décrire la méthode utilisée pour créer des données dérivées (traitements effectués, logiciels (y compris le numéro de version) utilisés, analyses appliquées aux données).
- Décrire comment citer l'ensemble de données dans la documentation.  
DataCite.org recommande le format suivant : Créateur (Année de publication) : Titre. Version (facultatif). Éditeur. ResourceType (facultatif). Identifiant.

# Sommaire de section

- 1 Avant-propos
- 2 Introduction générale
- 3 Structure : sections et éléments
- 4 Exemples
- 5 Bonnes pratiques
- 6 Exercices**
  - 1. Éléments du PGD
  - 2. Bonnes pratiques
  - 3. Entrepôts de données
  - 4. Outil en ligne
- 7 Bibliographie
- 8 Annexe

## Exercice 1 : éléments du PGD

Trouver et décrire brièvement 10 principaux composants (informations sur les données, métadonnées de contenu, politiques d'accès et réutilisation...) qui seraient pertinents d'intégrer au PGD.

### Éléments-Candidats

Éléments	Description	Section

## Exemple de réponse : exercice 1

## Éléments-Candidats

Éléments	Description	Section
Dénomination du projet	nom, acronyme, identifiant	1
Agence de financement	type, nom, identifiant	
Responsabilités	responsable de chaque étape de la gestion des données (administrateur/gestionnaire de base de données), partenaires	2
Matériel	outils de gestion, sélection, nettoyage, normalisation et enrichissement des données	3



## Réponse : exercice 1 (suite)

### Autres éléments - section 1 - Informations sur le projet

- ~~de la convention de financement~~ ; Programme de recherche ; ~~Acronyme du projet~~ ; ~~Titre du projet~~ ; Objectifs du projet ; Mots-clefs du projet ; Coordinateur-Bénéficiaire ; Responsable scientifique du projet ; Affiliation et unité de rattachement du responsable scientifique ; Document référence

### Autres éléments - section 2 - Responsabilité des données

- ~~Nom du responsable de la gestion des don~~ au cours du projet de recherche, Propriété des données

### Autres éléments - section 3 - Ressources nécessaires à la mise en œuvre du PGD

- ~~Mai~~ ; Personnel ; Formations ; Montant financier

## Réponse : exercice 1 (suite)

### Autres éléments - section 4 - Jeux de données

- Nombre de jeu(x) de données ; Identifiant et nom du jeu de données ; Nature des données ; Réutilisation de données existantes ; Méthode de production des données ; Formats des données ; Support (de stockage) des données ; Volumétrie prévisionnelle ; Stockage et enregistrement des données ; Type d'hébergement ; Risques ou menaces sur les données ; Garantie de confidentialité des données ; Garantie d'intégrité et de traçabilité ; Lecture des données ; Garantie de disponibilité des données ; Gestion des accès ; Echanges et partage ; Standards et formats disciplinaires des métadonnées ; Mode de production et responsabilité des métadonnées ; Arborescence de classement ; Règles de nommage des jeux de données ; Documentation associée ; Principe général de diffusion ; Type de licence ; Potentiel de réutilisation ; Existence de publications associées aux données ; Dépôt et diffusion des données ; Identification des jeux de données sensibles ; Justification du principe d'exception aux conditions générales de diffusion ; Mesures de protection ; Embargo

### Autres éléments - section 5 - Archivage

- Sort des données à l'issue du projet ; Sélection des données ; Volume final des données ; Durées de conservation préconisée ; Plateforme d'archivage

## Exercice 2 : bonnes pratiques

Ajouter à la liste de bonnes pratiques de nouvelles recommandations pour la rédaction du PGD.

### Documentation-Bonne pratiques

Intitulé	Description

## Réponse : exercice 2

### Documentation-Bonne pratiques

Intitulé	Description
Contenu du projet	Décrire le contexte scientifique. Pourquoi les données ont été recueillies (questions, hypothèses), conditions lors de la collecte (où et quand elles ont été collectées), et normes utilisés.
Identifiants uniques	Affecter des identifiants uniques aux données pour citer facilement les données à partir de publications (Handle.Net Registry (HNR)).
Métadonnées	Utiliser des schémas et des normes de métadonnées si possible pour que vos données soient décrites selon une procédure commune ou connue (DC, DDI, TEI, EAD, METS, OLAC, DIMDI).
Documentation	Décrire la base de données (modèle conceptuel, logique), les scripts, algorithmes.
Dictionnaire de données	Décrire les variables utilisées, leurs définitions, codages (texte, nombre entier), les coordonnées spatiales, les date/heure (HH :MM :SS ou HHMMSS)

### Outil pour nettoyer les données (enregistrements, variables)

- OpenRefine (ex. Google Refine). *Nettoyage des données*. Disponible à l'adresse : <http://openrefine.org> [consulté le 6.04.2017]

## Exercice 3 : entrepôts de données

Trouver un entrepôt de données dans sa discipline à partir du site re3data. Le résumer brièvement dans le tableau suivant.

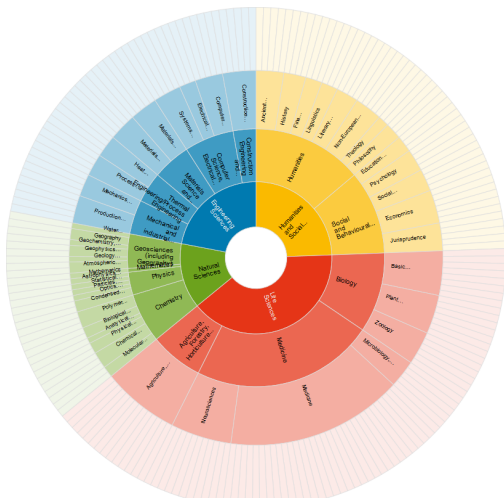
### Tableau à remplir

- nom de l'entrepôt
- sujets
- pays
- dépositaires admissibles
- propriété des données
- types de contenus
- types de métadonnées et sources
- langue
- licences

### Source

- Accès au site re3data : <http://www.re3data.org>

# re3data : moteur de recherche pour trouver un entrepôt



## Source

● re3data : <http://www.re3data.org>

## Exercice 3 : RESIF Seismic Data Portal

### portail RESIF

- sujets-mots clés : sismologie, expérimentation sur le terrain, volcan, sismomètre, océan, géophysique, géodésie, géosciences, gravimétrie, sciences naturelles
- pays : tout territoire
- dépositaires admissibles : ?
- propriété : Tous les contenus téléchargés restent la propriété des personnes ou institutions mentionnées dans les métadonnées.
- types de contenus : données statistiques, données de la recherche, images, bases de données, données archivées, texte structuré, documents bureautique
- types de métadonnées et sources :
- langue : français, anglais
- licences : Si vous utilisez des données issues de RESIF dans l'une de vos publications, vous devez faire apparaître la mention suivante : "*<RESIF est un équipement national, inscrit sur la feuille de route française des Infrastructures de Recherche éditée par le ministère de la Recherche...>*"

### Source

- Accès aux données RESIF : <http://www.resif.fr/spip.php?article26>

## Exercice 4 : outil en ligne OPIDoR

Utiliser en ligne l'outil OPIDoR.

### Consignes

- créer un compte
- choisir un modèle de PGD (Funder, Commission Européenne)
- partager le plan avec un collègue (shared)
- exporter le plan (pdf)

### Source

- DMP OPIDoR : <https://dmp.opidor.fr>



# Conclusion

Produire un PGD est essentiel pour un projet européen (supporter le cycle de vie des données en fournissant des services pour le stockage, la découverte et l'accès permanent)

## Car un PGD contient des informations

- scientifiques (informations de base sur la recherche)
- méthodologiques (comment sont organisées les données et la documentation les décrivant, à qui sont partagées les données)
- techniques (où sont stockées les données numériques et les sauvegardes)
- juridiques (copyright, propriété intellectuelle, licence de diffusion)
- **Ces contenus sont nécessaires à la manipulation de grandes quantités de données (*BigData*).**

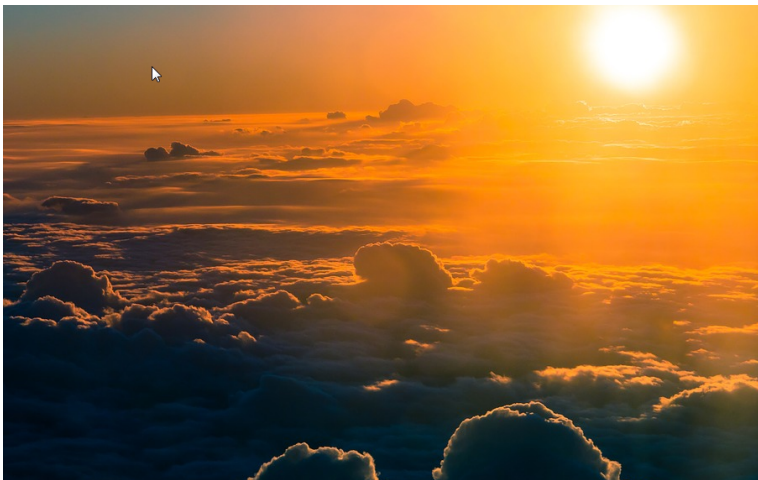
## Recommandation

- Le moment est (peut être) venu de s'impliquer dans la création d'un PGD
  - en utilisant éventuellement un outil centralisée (DMP OPIDoR).

## Localement

- Il me semble utile de relever ce défi de rédaction de PGD pour :
  - (au moins) garder une trace d'un projet de recherche dans un laboratoire (tant que les données ne sont pas perdues)
  - (mieux) se projeter plus facilement vers de nouveaux projets de recherche

# Illustration



: <https://pixabay.com/fr/users/xegxef-1679630> :

# Sommaire de section

- 1 Avant-propos
- 2 Introduction générale
- 3 Structure : sections et éléments
- 4 Exemples
- 5 Bonnes pratiques
- 6 Exercices
- 7 Bibliographie**
- 8 Annexe

## Bibliographie

## 10 livres

- ACCART J.-P. RIVIER A.. *Mémento de l'information numérique*. Paris : Éd. du Cercle de la librairie, 2012. (Bibliothèques). ISBN : 9782765413325.
- BARRON G. LE GOFF-JANTON P. (Eds.). *Intégrer des ressources numériques dans les collections*. Villeurbanne : Presses de l'Enssib, 2014. 183. (Collection La Boîte à outils 29). ISBN : 9791091281188.
- BRAND-FROISSAC O., CHARDON L., DAVID M. [et al.]. *Guide de bonnes pratiques pour les administrateurs systèmes et réseaux*. Meudon : Mission Ressources et Compétences Technologiques [MRCT], 2012. 130. ISBN : 9782918701071.
- GREEN J. L. KING J. H.. *Guidelines for the development of a Project Data Management Plan (PDMP)* [en ligne]. Greenbelt : NASA Goddard Space Flight Center, 1988. Disponible sur : <<https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19920018913.pdf>> (Accessible le : 20.02.2017)
- LEE S. D.. *Building an electronic resource collection : a practical guide*. London : Library Association Publishing, 2002. 147. ISBN : 185604422X.
- MITCHELL E. T.. *Cloud-Based Services for Your Library*. Chicago : Amer Library Assn, 2013. 176. (Lita Guide). ISBN : 9781555708788.
- NG K. B. (Ed.). *Recent Developments in the Design, Construction, and Evaluation of Digital Libraries : Case Studies*. Hershey : IGI Global, 2013. 275. ISBN : 9781466629912.
- POLZIN-HAUMANN C. SCHWEICKARD W. (Ed.). *Manuel de linguistique française*. Berlin : De Gruyter, 2015. 729. (Manuals of Romance linguistics 8). ISBN : 9783110302080.
- PRYOR G., JONES S. WHYTE A. (Eds.). *Delivering research data management services : fundamentals of good practice*. London : Facet Publishing, 2014. 242. ISBN : 9781856049337.
- WITTEN I. H., BAINBRIDGE D. NICHOLS D. M.. *How to build a digital library*. Burlington : Morgan Kaufmann, 2010. 629. (The Morgan Kaufmann series in multimedia information and systems). ISBN : 9780123748577.

# Sommaire de section

- 1 Avant-propos
- 2 Introduction générale
- 3 Structure : sections et éléments
- 4 Exemples
- 5 Bonnes pratiques
- 6 Exercices
- 7 Bibliographie
- 8 Annexe**

# Outils d'aide à la création de PGD en France : DMP OPIDoR

## Caractéristique

- nouveau service de l'Inist/CNRS à destination de l'Enseignement Supérieur et de la Recherche
- proposition de différents modèles préconisés par les financeurs, ainsi que des guides et des exemples personnalisés

Home About Help

Welcome.

DMP OPIDoR (DMP pour une Optimisation du Partage et de l'interopérabilité des Données de la Recherche) will help French researchers to write Data Management Plans.

Videos on how to use DMP OPIDoR

Se connecter

Log in

Institutional Access

Individual Access

Email address \*

Password \*

This field is mandatory

Remember me

Log in

Forgot your password?

Create an account

New to DMP OPIDoR? Create an account today.

: page d'accueil DMP OPIDoR :

# Outils d'aide à la création de PGD ailleurs : DMP online

C'est un outil pour créer des plans personnalisés selon le contexte du financer de la recherche.

## Pour les chercheurs

- exemples de PGD
- guides et bonnes pratiques

The screenshot shows the DMPonline website interface. At the top left is the 'DMP ONLINE' logo. A navigation bar contains links for 'Home', 'About', 'Future plans', and 'Help'. Below the navigation bar, a 'Welcome' message reads: 'DMPonline has been developed by the Digital Curation Centre to help you write data management plans.' To the right of the welcome message is a 'Sign in' form with fields for 'Email address \*' and 'Password \*', a 'Remember me' checkbox, and a 'Sign in' button. Below the sign-in form is a 'Sign up' section with the text 'New to DMPonline? Sign up today'. At the bottom of the page, there are logos for 'DCC' and 'Jisc', and contact information: 'Contact us | Terms of use | DMPonline previous version © 2004 - 2017 Digital Curation Centre (DCC)'.

: page d'accueil DMP online :

milesker  
gracias  
merci