

Introduction aux plans de gestion et de partage de données



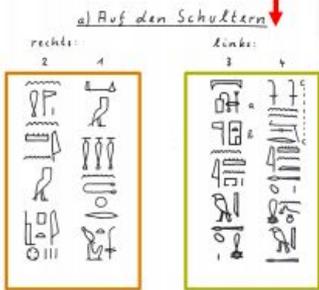
Emmanuelle Morlock -IE CNRS
Laboratoire Histoire et sources des mondes antiques - UMR 5189

Qui parle ?

Soutien à l'édition critique de sources anciennes en TEI



Block statue of Hr (III), son of Nsr-Jmn. Cairo CG 42230
Upper side – texts on right and left shoulders



Front side

/TEI/sourceDoc

```
<surface xml:id="ck68-upperSide" ana="front-side">Face supérieure de la statue</surface>
```

Text layouts

/TEI/teiHeader/fileDesc/sourceDesc/msDesc/
physDesc/objectDesc/layoutDesc

```
<layout xml:id="ck68-rightShoulder" columns="2" corresp="#ck68-upperSide">
<desc>Deux colonnes de texte sur l'épaule droite</desc></layout>
<layout xml:id="ck68-leftShoulder" columns="2" corresp="#ck68-upperSide">
<desc>Deux colonnes de texte sur l'épaule gauche</desc></layout>
```

Abstract text

/TEI/teiHeader/fileDesc/sourceDesc/msDesc/msContents/msItem

```
<msItem xml:id="ck68-text3" ana="meta-egypt-ths:royal-donation-formula">
<title type="modern">Formule de donation royale</title>
<textLang mainLang="egy-x-egt"/>
</msItem>
```

Text edition

/TEI/text/body/div[@type="edition"]

```
<div xml:id="ck68-shouldersTranslit" type="textpart"
corresp="#ck68-rightShoulder #ck68-leftShoulder">
<ab corresp="#ck68-text3">
<milestone unit="layout" corresp="#ck68-rightShoulder"/>
<lb n="1" rend="verti-right-to-left"/>dy m ḥsw.t n.l ḥr nswt
<lb n="2" rend="verti-right-to-left"/>n hm-nṛ n lmn m ḥp.t-sw.t
<milestone unit="layout" corresp="#ck68-leftShoulder"/>
<lb n="3" rend="verti-left-to-right"/>hry sšw ḥw.t-nṛ n pr lmn Hr
m3' ḥrw s3 mi n
<lb n="4" rend="verti-left-to-right"/>Ny-se-r-lmn m3' ḥrw s3 mi
nw Hr m3' ḥrw
</ab>
</div>
```

Campus Condorcet



Sommaire

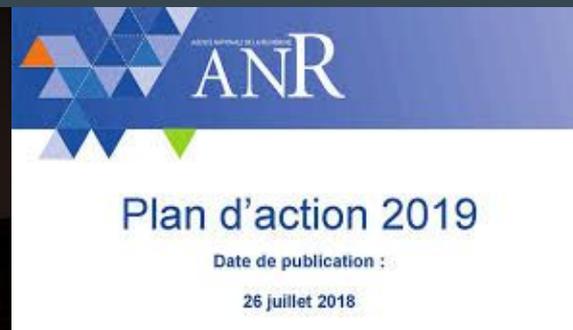
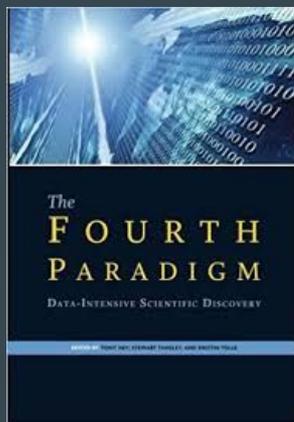
- Le contexte : "déluge" des données numériques et Science ouverte
- Définir les données de la recherche ? pas si facile...
- De nouvelles pratiques de gestion...
- ... à mettre en oeuvre tout au long du cycle de vie des données

- Fonction et contenus des plans de gestion et de partage de données
- Des principes "FAIR" pour guider la gestion des données et la rédaction des plans
- Workflow et calendrier
- Types et modèles
- La combinaison "dépôt + data paper"
- Exemples

- Quelques outils d'aide à la rédaction
- la question de l'archivage pérenne
- Portail Doranum pour s'auto-former

Contexte

Généralisation du numérique & politiques d'accès ouvert



“Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications.””

Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014

Données ?

Données de la recherche : une notion à “géométrie variable”

- Souvent définies comme des “enregistrement factuels” (chiffres, textes, images et sons) utilisés comme sources principales de la recherche et nécessaires à la validation des résultats...
- ... issues de l'observation, de l'expérimentation ou dérivées de sources existantes...
- mais il reste impossible de les définir dans l'absolu hors contexte scientifique

Une grande variété de d'objets et de formats

Algorithmes
Modèles
Scripts

Méthodologies
Workflows

Questionnaires
Transcriptions
Carnets de notes

Procédures
Protocoles

Fichiers audio et vidéo

Documents texte
électroniques
(pdf, doc, txt)

Objets numériques (figures, vidéos)

Autres fichiers
(revue de la littérature,
emails)

Carnets de laboratoire,
Carnets de terrain,
Journaux

Schémas de BDD

feuilles de calcul (xls)

Collections
de ressources
numériques

Résultats
de tests
d'expériences

Logiciels : fichiers de
configuration

Photographies
Films

Contenus d'une BDD
(audio, vidéo, texte, images)

Périmètres habituels

Toutes les données et métadonnées :

- générées,
- collectées,
- modifiées,
- ou enrichies

au cours d'une recherche

- Cahiers de labo
- Analyses préliminaires
- Projets de documents
- Examens par les pairs
- Communications personnelles
- Objets matériels

- Données administratives
- Supports de formation
- Publications

Spécificité des données en SHS

- Données souvent **qualitatives** , grande **hétérogénéité**
- Pas des preuves en soi mais des **objets multi-dimensionnels mobilisés comme preuves** à l'appui d'une argumentation... (Trevor Owens).
- Elles ne possèdent **pas caractéristiques propres** qui permettent de les définir de manière absolue et indépendamment du contexte du travail scientifique (Schöpfel, Kergosian et Prost).

**“A digital, selectively
constructed, machine-actionable
abstraction representing some
aspects of a given object of
humanistic inquiry.”**

Schöch, Christoph. “Big? Smart? Clean? Messy?” Data in the Humanities”.
Journal of Digital Humanities. Vol. 2, n°3, 2013

Alexandre Serres, Marie-Laure Malingre, Morgane Mignon, Cécile Pierre, Didier Collet.
Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs :
une enquête à l'Université Rennes 2 [Rapport de recherche] Université Rennes 2. 2017,

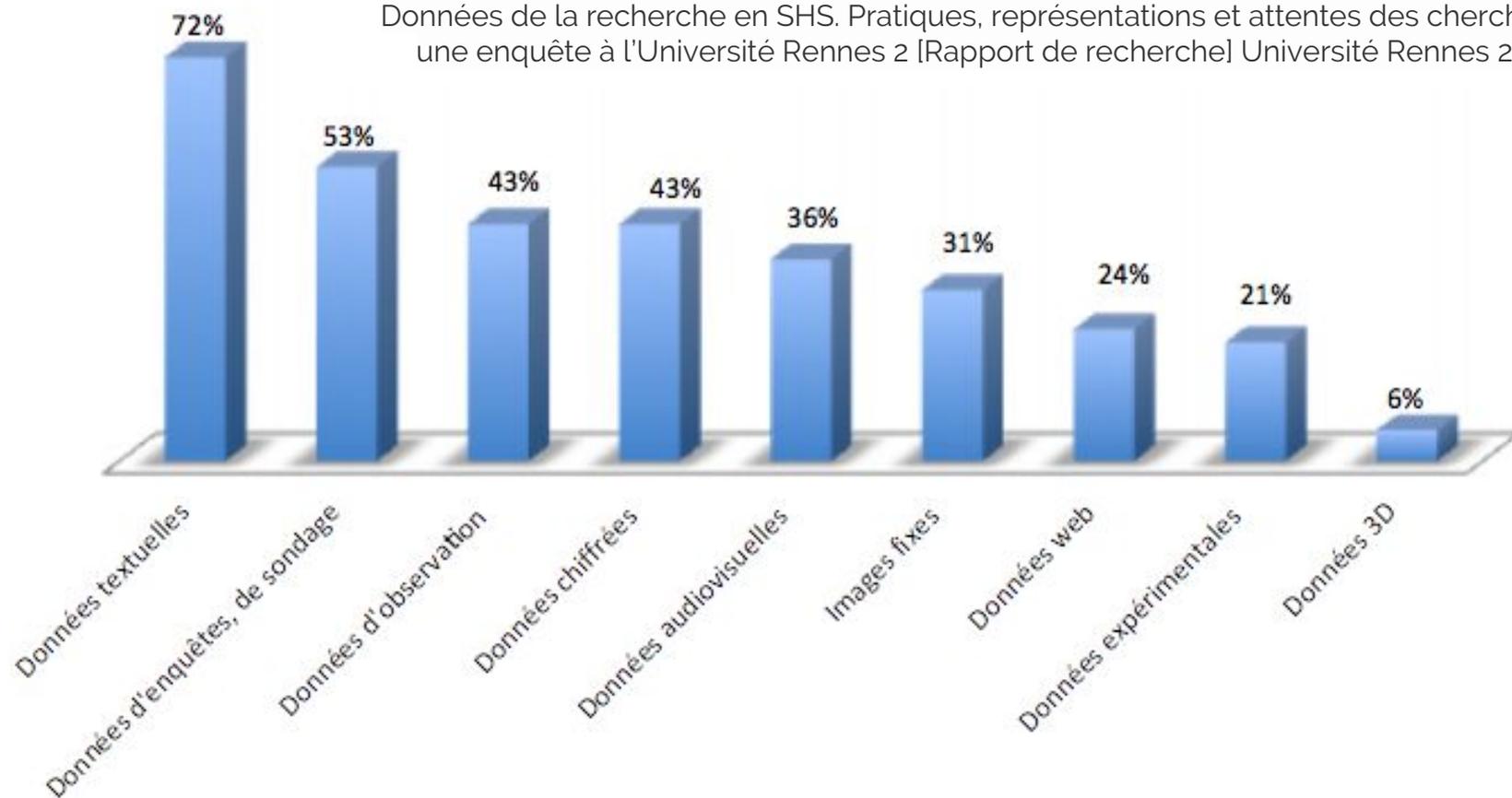


Figure 13 : Catégories des données sources (N = 127)

Informatique & droit

Vocabulaire de l'informatique

- “Représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement” (*Arrêté sur l'enrichissement du vocabulaire de l'informatique de 1981*)

Droit

- Pas de définition juridique !
- 3 domaines :
 - propriété intellectuelle
 - données publiques (open data)
 - données personnelles

Agnes Robin. “Les données de la recherche scientifique et le droit”.

MSH SUD, 16 octobre 2018

<https://www.youtube.com/watch?v=nJx9r4qP-A>

“Comment bien perdre ses données ?”*



“Comment bien perdre ses données ?”



- Vous ne savez pas comment relire une cartouche Iomega JAZ.
- Votre disquette 3 1/2 pouces est endommagée.
- Vous avez effacé par mégarde le dossier où était votre fichier.
- Le codec utilisé pour votre image n'est plus reconnu par aucun de vos outils.
- Vous ne savez plus comment se nomme le fichier ni sur quel disque vous l'avez mis .
- Vous avez 40 exemplaires d'un même fichier mais vous ne savez pas lequel est le bon.
- La lecture de votre fichier par la nouvelle version du logiciel vous donne un autre résultat qu'auparavant.
- Vous ne pouvez plus installer votre unique outil de visualisation sur la nouvelle version de votre système d'exploitation.
- Vous revenez 10 ans plus tard sur un de vos fichiers, mais vous ne comprenez plus sa structure ni son contenu que vous n'avez jamais documenté.
- Etc.

Extrait d'une présentation de la
TGIR Huma-num (juin 2018)

https://rhn2018.sciencesconf.org/data/20180613_archivage.pdf

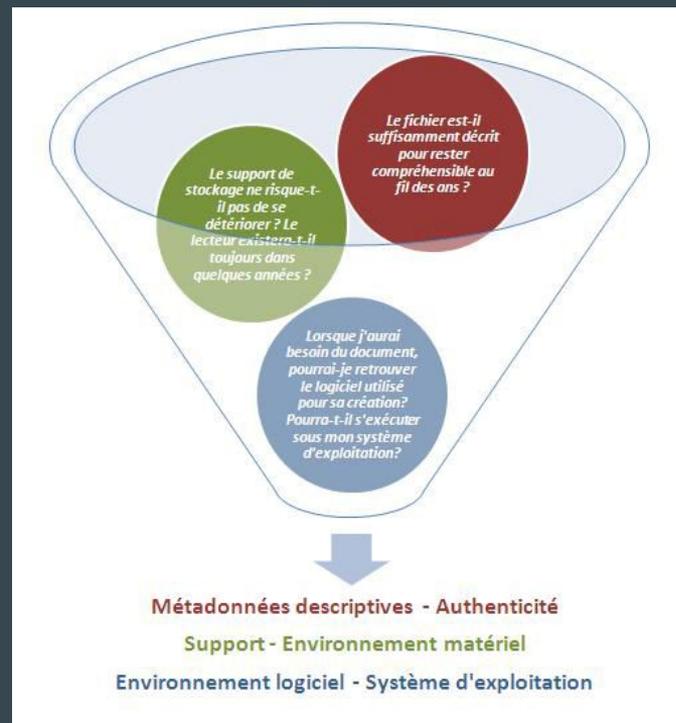
Les données numériques ont besoin d'une gestion active

Les 5 risques principaux

- obsolescence
- vieillissement des supports
- mauvaises manipulations
- manque d'organisation
- perte de la signification du contenu (documentation)

Les solutions

- copies et technologies de stockage multiples
- veille technologique
- formats ouverts, standards et des normes
- conserver les métadonnées avec le document



source : CINES

<https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/>

Bénéfices de la gestion des données de la recherche

Pour le chercheur

- Mieux gérer les données qu'il produit ou collecte
- S'inscrire dans le plan national de la Science Ouverte (structurer et ouvrir ses données selon les principes FAIR)
- Obtenir des financements
- Mieux promouvoir ses recherches

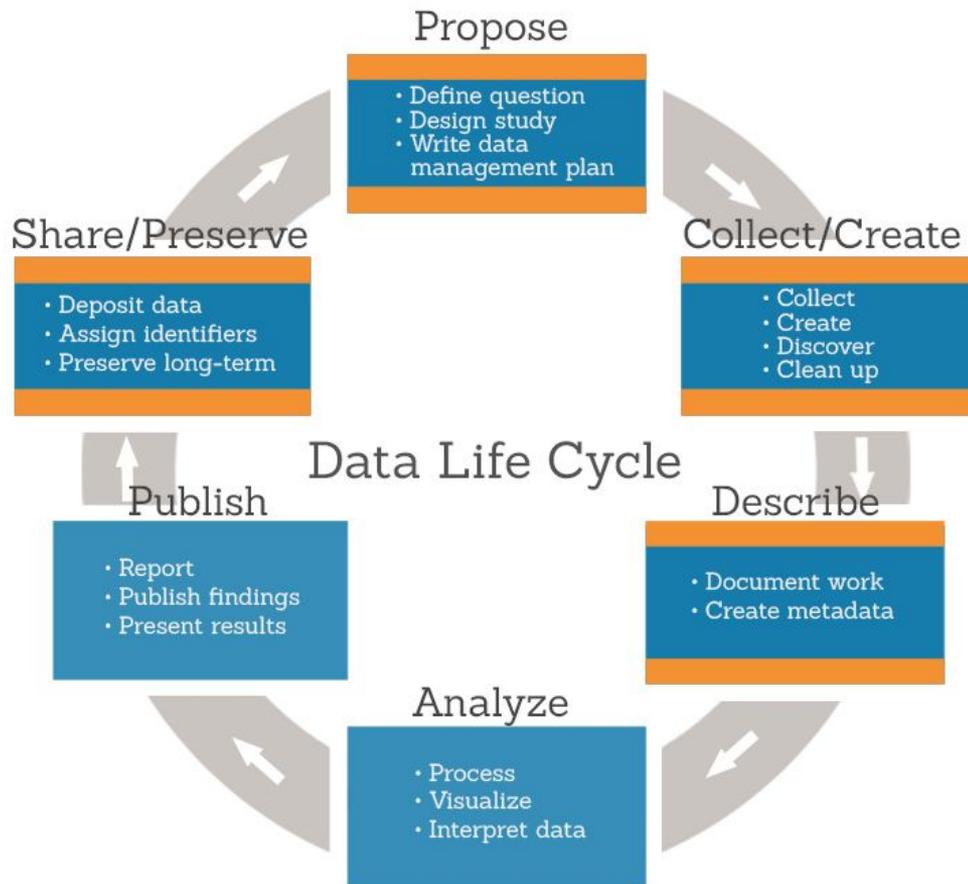
Pour la science

- Ouverture et transparence garantie contre la fraude scientifique
- Les bonnes pratiques de gestion ont un impact sur la qualité des recherches
- Rentabiliser l'investissement dans la recherche en facilitant les réutilisations
- Permettre de nouvelles recherches (fouille de données, lecture distante / data driven science, visualisations synthétiques pour produire de nouvelles connaissances)

Pour la société

- Les recherches bénéficient à tous, renforcent la confiance des citoyens et leur permet de s'impliquer
- Éclairer la société sur ses évolutions les plus récentes via l'open data et le big data

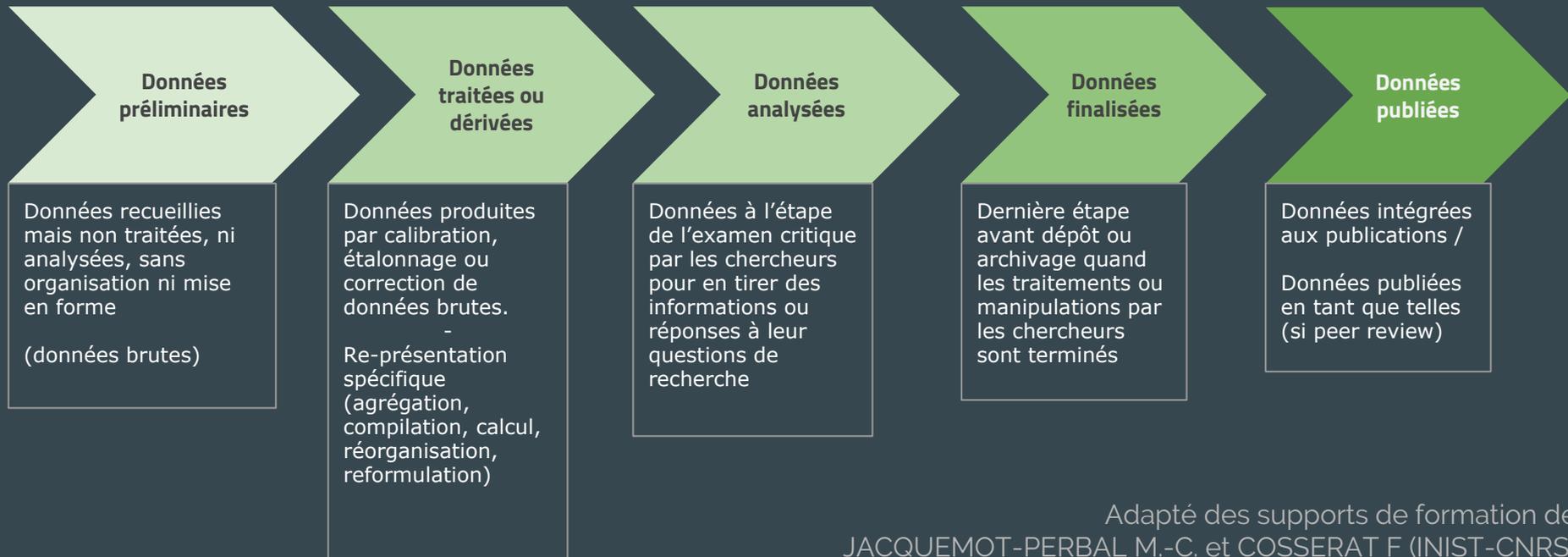
Tout au long des du cycle de vie



Source : Bibliothèque de l'Université de
Californie San Diego

<https://library.ucsd.edu/lpw-staging/research-and-collections/data-curation/data-management/index.html>

Au fil du cycle de vie : un continuum d'états



Adapté des supports de formation de
JACQUEMOT-PERBAL M.-C. et COSSERAT F (INIST-CNRS)

La solution : gérer les données de la recherche et planifier cette gestion



Définition (Agence nationale de la Recherche)

“Document formalisé, établi dès le démarrage d'un projet ou processus de recherche, le plan de gestion des données définit comment les données seront créées/collectées et la manière dont elles seront documentées, utilisées, gérées, partagées, protégées et conservées au cours et à l'issue du projet. Ce document sera, jusqu'à l'achèvement du projet, continuellement mis à jour.”

Principes FAIR

F
Findable



A
Accessible



I
Interoperable



R
Reusable



F**Facilement trouvables**

- Comment va-t-on découvrir vos données ?
- Que faire pour les rendre compréhensibles en dehors de l'équipe du projet ?

A**Accessibles**

- Où va-t-on trouver vos données ?
- Sont-elles librement accessibles ?

I**Interopérables**

- Est-ce que les formats et les normes qu'elles utilisent leur permettent-elles d'être exploitées dans un écosystème numérique ?

R**Réutilisables**

- Les données sont-elles compréhensibles ?
- Les données sont-elles ouvertes
- Y-a-t-il des restrictions d'usage

F**Facilement trouvables**

- Métadonnées descriptives riches
- Identifiants pérennes
- Convention de nommage, versionnage
- Mots clés pour les moteurs de recherche

A**Accessibles**

- Quelles modalités d'accès ?
- Techniques, logiciels, protocoles pour accéder aux données
- Où sont déposés les codes, la documentation ?

I**Interopérables**

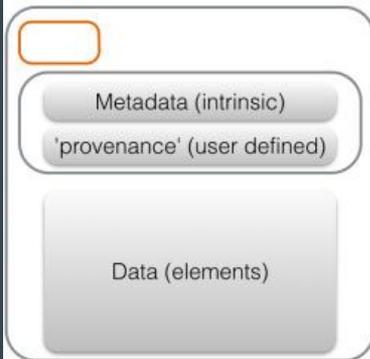
- Est-ce que les formats et les normes qu'elles utilisent leur permettent-elles d'être exploitées dans un écosystème numérique ?

R**Réutilisables**

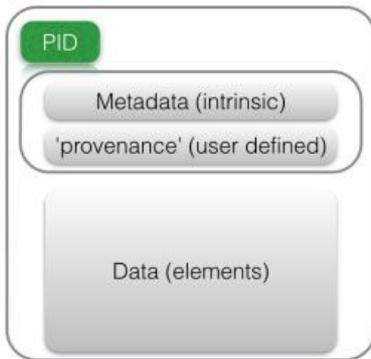
- Les données sont-elles compréhensibles ?
- Les données sont-elles ouvertes ?
- Y-a-t-il des restrictions d'usage ?

Data as increasingly FAIR Digital Objects

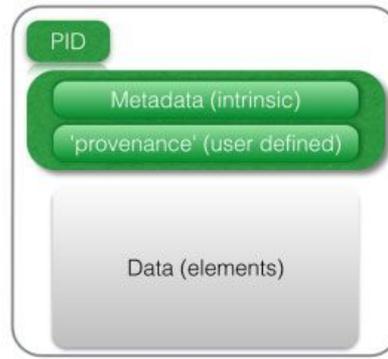
Totally UNFAIR



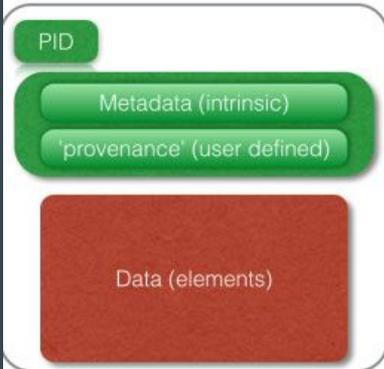
Findable
Usable for Humans



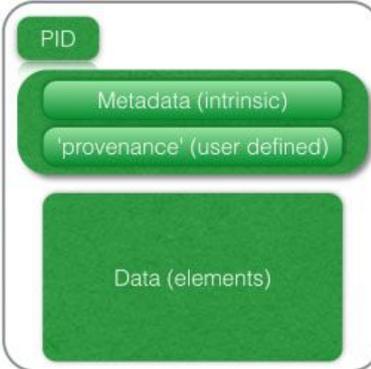
FAIR metadata



FAIR data-
restricted access



FAIR data-
Open Access



FAIR data-
Open Access/Functionally Linked



Processus et calendriers

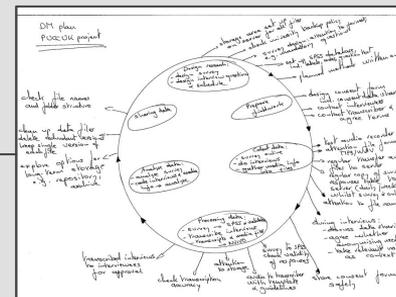


Inventaires questions audits

Politiques

Consultations

Data Stage	Output	# of Files / Typical Size	Format	Other / Notes
Primary Data				
Raw				
Processed				
Analyzed				
Finalized				
Ancillary Data				
Ancillary Data #1				
Ancillary Data #2				



Choix d'un modèle

DCC, ERC, Univ. paris diderot, Inrap...

Saisie dans un outil collaboratif



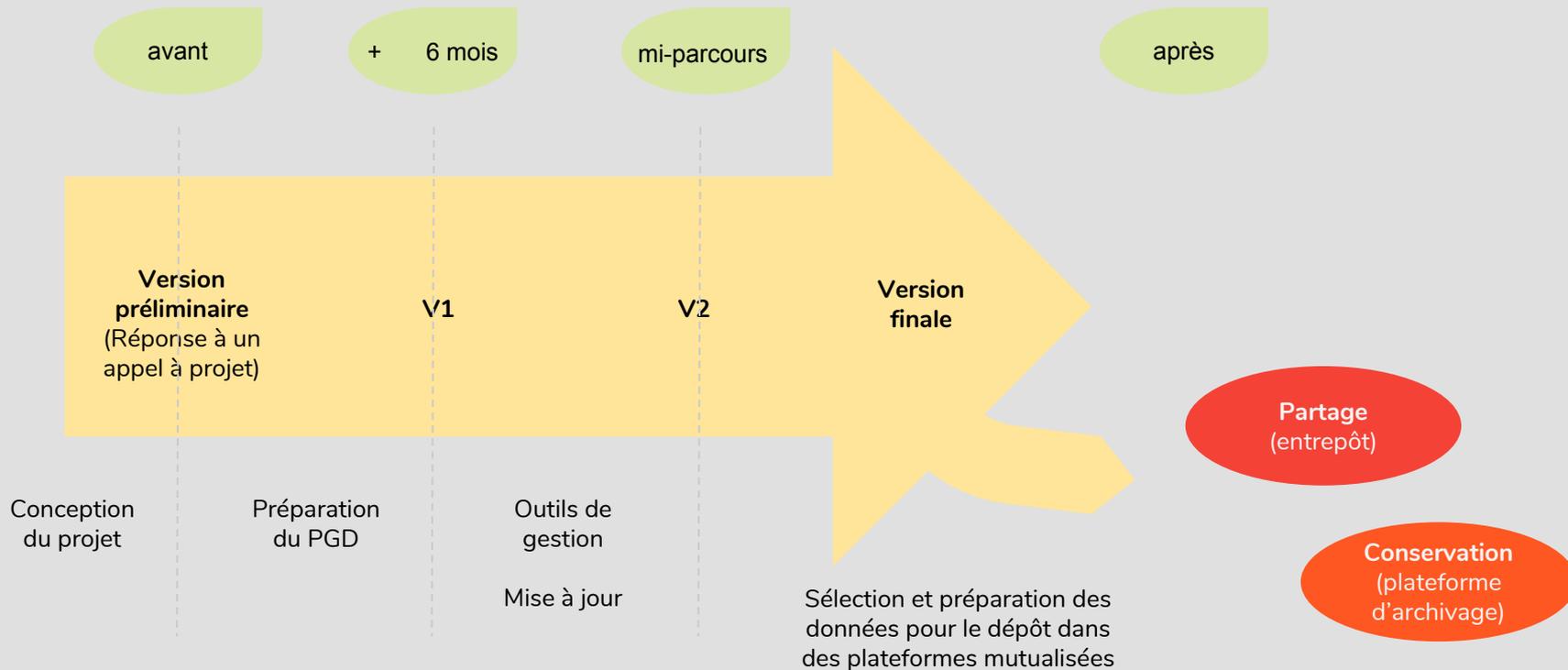
Outils à développer

Validation technique

Révisions

Rendu final

Calendrier (H2020)



Plusieurs types de plans...

- Le **document prospectif** initial (½ à 4 pages, selon les contraintes du financeur)
- Le **document évolutif** pendant le projet (dizaines de pages)
- Le **document finalisé** et livré au financeur (id.)
- puis reprise des infos dans les formulaires de dépôts des plateformes et dans des **data papers** (structurés selon les consignes de la revue)

Modèles-types à connaître

- Modèle du Digital Curation Centre (RU)
- Modèles H2020 - ERC (démontrer l'état de FAIRisation des données)
- Modèle de Paris Descartes / Diderot
- Modèles institutionnels (INRA, IRSTEA...)
- Modèle pour les logiciels : PRESOFT



[Renseignements sur le projet](#)[Vue d'ensemble](#)[Contexte, acteurs, ressources](#)[Stockage, partage, protection et diffusion](#)[Dissémination et archivage](#)[Partager](#)[Télécharger](#)

FAIR DMP Paris Descartes/Diderot

Ce plan est basé sur le modèle "FAIR DMP Paris Descartes/Diderot" fourni par Université de Paris 7 - Denis Diderot.

Ce plan de gestion des données est **proposé par les Universités Paris Descartes et Paris Diderot**. Il s'adresse prioritairement aux chercheurs affiliés à ces deux établissements et à la ComUE Sorbonne Paris Cité mais reste un outil accessible et utilisable par tous. Il **intègre dans sa structure et ses recommandations les principes des FAIR Data** désormais attendus de certains financeurs comme la Commission Européenne. Si le plan de gestion de données doit rester **un document "unique"**, nous proposons une structuration en **4 étapes** fruit de nos échanges avec le terrain. Les trois premiers onglets peuvent faire l'objet d'une **première version adressée dans les 6 premiers mois** suivant l'acceptation du projet (critère H2020) puis de mises à jour régulières. L'**onglet 4** étant, quant à lui, **attendu lors de la phase finale d'évaluation** du projet. Au regard de notre expérience, ce découpage nous semble plus réaliste vis à vis de la montée en charge d'un projet de recherche et plus à même de garantir une gestion des données conforme aux objectifs du projet.

[Contexte, acteurs, ressources \(3 sections, 12 questions\)](#)[Stockage, partage, protection et diffusion \(4 sections, 18 questions\)](#)[Dissémination et archivage \(5 sections, 9 questions\)](#)

**La combinaison gagnante
dépôt + data paper**

Reading: Archaeological Survey and Monitoring Data from the Flower of Ugie, Wrecked 1852 in the Easte...

Share: [f](#) [t](#) [g+](#) [in](#)

Download [A-](#) [A+](#) Alt. Display

Data papers

Archaeological Survey and Monitoring Data from the Flower of Ugie, Wrecked 1852 in the Eastern Solent, England

Author: [Julian Whitewright](#)

Abstract

This paper discusses the potential future reuse of the archived dataset resulting from the archaeological survey of the shipwreck remains of the Flower of Ugie, a wooden sailing barque wrecked in the Eastern Solent in 1852. Seven years of archaeological survey are represented in the site plans which record the structural details of the vessel along with many of its material characteristics. This record serves to establish the baseline condition of the site, against which future monitoring of the site will be conducted allowing informed management decisions to be made.

Keywords: [aggregate](#) , [archaeology](#) , [Flower of Ugie](#) , [maritime](#) , [shipwreck](#)

JUMP TO

DISCUSSIONS

[Abstract](#)[Context](#)[Methods](#)[Dataset Description](#)[Reuse Potential](#)[Acknowledgements](#)[References](#)



Index

Auteurs

Mots-clés

Index géographique de
référence

Années

Langues

Anniversaire

Les 20 ans de Cybergeo

Rubriques

Data papers



2019

892

A harmonized database to follow the demographic trajectories of European cities, the TRADEVE database (Demographic Trajectories of European Cities)

Une base de données harmonisée pour suivre les trajectoires démographiques des villes européennes, la base TRADEVE (Trajectoires Démographiques des Villes Européennes)

Marianne Guérois, Anne Bretagnolle, Antonin Pavard, Paul Gourdon et Natalia Zdanowska

Plan

Language

Theme

Spatial coverage

Time coverage

Format name and version

File's format

Creation date

Dataset creator

Name and function developed by the person responsible for the resource

Responsible organization and person

Repository location

Licence

Use

Reuse potential

Type of spatial representation

Source

Format for metadata

Acknowledgements

Le dépôt

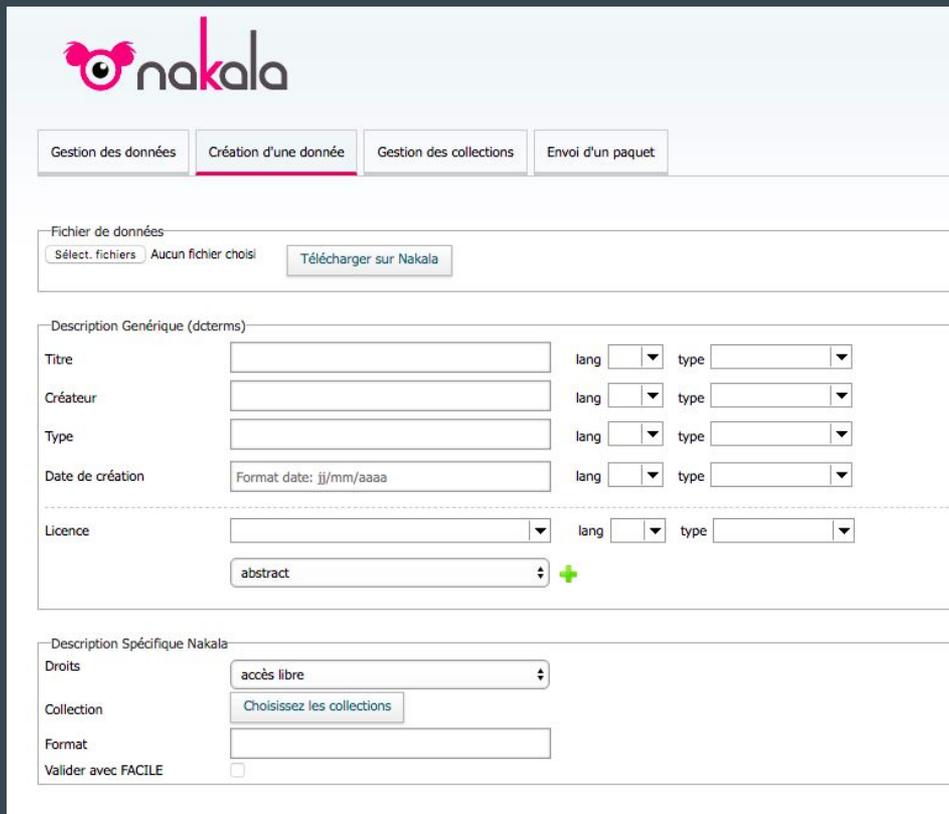
répondre aux exigences des financeurs,
augmenter la visibilité,
permettre la validation

Critères de choix d'un entrepôt

Si le choix de l'entrepôt ne vous est pas imposé par le financeur, l'institution ou la revue, quelques questions à se poser :

- disciplinaire ?
- certifié ?
- données hébergées en France ?
- archivage à long terme ?
- coûts ?
- types de données acceptées ?
- formats ?
- modifications possibles ?
- attribution d'un identifiant pérenne ?
- licence imposée ?
- restrictions d'accès ?
- embargos ?
- statistiques d'utilisation ?

Nakala : identifiant pérenne, accès persistant, sécurisation



The screenshot shows the Nakala web interface. At the top left is the Nakala logo. Below it are four navigation buttons: 'Gestion des données', 'Création d'une donnée' (highlighted with a red underline), 'Gestion des collections', and 'Envoi d'un paquet'. The main content area is titled 'Fichier de données' and contains a 'Sélect. fichiers' button, the text 'Aucun fichier choisi', and a 'Télécharger sur Nakala' button. Below this is the 'Description Générique (dcterms)' section, which includes fields for 'Titre', 'Créateur', 'Type', and 'Date de création', each with a text input and 'lang' and 'type' dropdown menus. The 'Date de création' field has a 'Format date: jj/mm/aaaa' hint. Below this is the 'Licence' section with a dropdown menu and 'lang' and 'type' dropdowns. A dropdown menu is open showing 'abstract' with a green plus sign. The bottom section is 'Description Spécifique Nakala', containing a 'Droits' dropdown menu with 'accès libre' selected, a 'Collection' button labeled 'Choisissez les collections', a 'Format' text input, and a 'Valider avec FACILE' checkbox.



**Préparation de
l'archivage à long
terme**

QUploadCommunitiesemmanuelle.morlock@morn.fr

DeleteSavePublish

New upload

instructions: (i) Upload minimum one file or fill-in required fields (marked with a red star). (ii) Press "Save" to save your upload for editing later. (iii) When ready, press "Publish" to finalize and make your upload public.

Files Choose files Start upload

Drag and drop files here

— or —

Choose files

(minimum 1 file required, max 50 GB per dataset - contact us for larger datasets)

Communities recommended

Start typing a community name... Q

Upload type required

PublicationPosterPresentationDatasetImageVideo/AudioSoftwareLessonOther

 Digital Object Identifier

Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.

 Publication date *

Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.

 Title *

Required.

 Authors *  

Optional.

[+ Add another author](#)

 Description *

  **B** *I* **S** ~~x~~ ^x         *I* Σ Ω  Source 

Required.

 Version

Optional. Mostly relevant for software and dataset uploads. Any string will be accepted, but semantically-versioned tag is recommended. See [Zenodo.org](#) for more information on semantic versioning.

New upload

Instructions: (i) Upload minimum one file or fill-in required fields (marked with a red star). (ii) Press "Save" to save your upload for editing later. (iii) When ready, press "Publish" to finalize and make your upload public.

Files >

 Choose files

 Start upload

Communities 

recommended >

Upload type

required >

Basic information

required >

License

required >

Funding

recommended >

Related/alternate identifiers

recommended >

Contributors

optional >

References

optional >

Journal

optional >

Conference

optional >

Book/Report/Chapter

optional >

Thesis

optional >

Subjects

optional >

 Delete

 Save

 Publish

Exemples

Dataset summary

- Lister les jeux de données (nom + référence interne)
- Origine, type, volume
- Types de données / formats
- *Potentiel de réutilisation : à qui seront-elles utiles ?*

Table 5.1 – Potential datasets

Data Type	Format	Volume	IPR Owner
Work Package 2 - D2.1: Report on preparation of nanotemplates for mAb crystallization (lead: IMP)			
Experimental data – Brunauer–Emmett–Teller (BET)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – scanning electron microscope (SEM)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – transmission electron microscopy (TEM)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – crystal shape/size measurement	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – X-ray diffraction (XRD)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – High performance liquid chromatography (HPLC)	XLSX, JPEG, PDF	< 100 MB	IMP
Experimental data – dynamic light scattering (DLS)	XLSX, JPEG, PDF	< 100 MB	IMP
Synthesis protocol	DOC, PDF, JPEG	< 100 MB	IMP
Lab notes	DOC, PDF	< 100 MB	IMP

Makind data “Findable”

Quelles sont les mesures prises pour rendre les données “trouvables” et compréhensibles ?

- Attribution d’identifiants pérennes aux (meta)données
 - Métadonnées descriptives riches, lisibles par les machines
 - Documentation sur la manière dont sont collectées, générées, organisées les données
 - Bien identifier les versions avec date et numéro de version, garder la version originale
 - Les métadonnées incluent l’identifiant de la donnée qu’elles décrivent
 - Les (méta)données sont référençables et indexables par les moteurs de recherche / moissonneurs (mots clés définis selon des référentiels connus et utilisés par la communauté cible)
-
- Préciser comment est obtenu l’identifiant pérenne (service dédié, plateforme de dépôt)
 - S’assurer que le lien entre données et métadonnées via l’identifiant est bien assuré
 - Convention de nommage
 - Identification des versions par date (20180919) et numéro de version (V01)
 - Dire s’il existera un fichier de documentation de type “readme” (par ex. acronymes, dictionnaire de données, mention des méthodes, techniques...)

Example

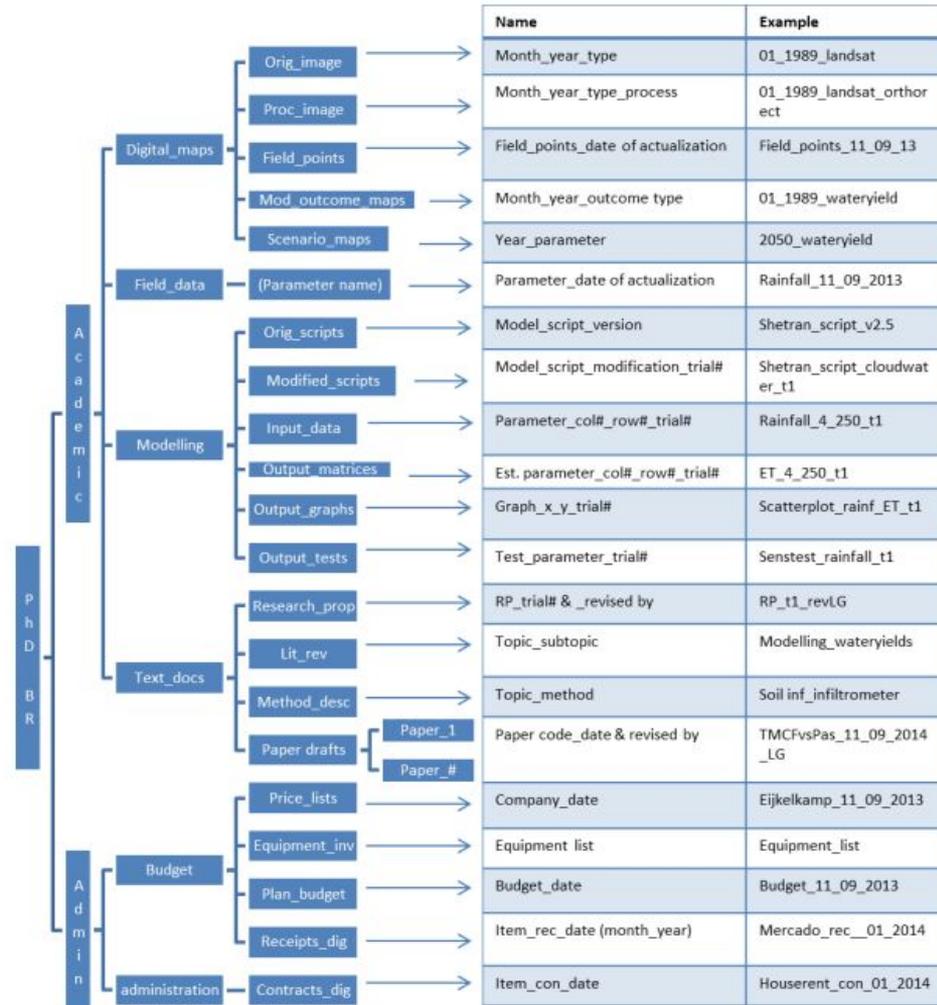
A DOI will be assigned to datasets for effective and persistent citation when it is uploaded to the repository [Zenodo]. This DOI can be used in any relevant publications to direct readers to the underlying dataset.

Each dataset generated during the project will be recorded in an Excel spreadsheet with a standard format and allocated a dataset identifier, see tables 6.1.1 and 6.1.2 below. The spreadsheet will be hosted at UST [ShareFile]. This dataset information will be included in the metadata file (see Section 6.3 and Annex B) at the beginning of the documentation, and updated with each version.

Exemple

AMECRYS naming convention for project datasets will comprise of the following:

1. A unique chronological number of the datasets in the project will be added.
2. The title of the dataset.
3. For each new version of a dataset it will be allocated with a version number which will be for example start at v1.0.
4. A prefix "AM" indicating an AMECRYS dataset.
5. A unique identification number linking with the dataset work package and deliverable/task e.g., "W4_D4.3". 01_Crystallization Kinetics_v1.0._AM_W4_D4.3.xlsx



Making data “openly accessible”

Conditions d'accès aux données

- Quelles données ? Bien identifier les données librement mise à disposition
 - Où sont déposées les données, les codes et la documentation (quels entrepôts) ?
 - Comment ? Expliquer comment on accède aux données (autorisations, protocoles, méthodes...)
-
- Toute personne doté d'un ordinateur et d'une connexion internet peut au moins accéder aux métadonnées
 - S'il y a besoin de méthodes ou de logiciels nécessaires pour accéder aux données, bien les préciser et fournir la documentation associée, dire où trouver les codes source...
 - Bien justifier les non mises à disposition et les restrictions d'accès (par ex. accès sur authentification pour les membres du projet)
 - Expliquer comment les métadonnées resteront accessibles, même quand les données ne le sont pas / plus

Example

Research data should be deposited in the Zenodo repository as soon as possible unless a decision has been taken to protect results. Specifically, research data needed to validate the results in the scientific publications should be deposited in the data repository at the same time as publication (GA Art. 29.3).

During embargo periods, information about the restricted data will be published in the data repository, and details of when the data will become available will be included in the metadata. Where a restriction on open access to data is necessary, attempts will be made to make data available under controlled conditions to other individual researchers.

There will be three restricted datasets within deliverable D6.3 (lead: CPI):

- Anti-CD20 CHO cell line
- HEL4 E. coli strain
- Growth protocols (although basic data such as final product titre would be available).

Making data “interoperable”

Des données prêtes pour l'échange, l'interprétation, la combinaison avec d'autres données d'une manière automatique ou semi-automatisée.

- Méthodes
 - Standards
 - Vocabulaires contrôlés, communs à une communauté de recherche / discipline
-
- Montrer comment on va produire des (meta)données lisibles et manipulables par les machines, c'est-à-dire formalisées dans un langage bien établi, lui aussi documenté
 - Les vocabulaires utilisés sont eux-même FAIR
 - Montrer comment les données du projet se lient automatiquement aux données d'autres projets (par ex. dans le web de données liées / *Linked Open Data*)

Increase data re-use

Mesures prise pour permettre et accroître la ré-utilisabilité des données

- Quelles données sont susceptibles d'être réutilisées ? pour combien de temps
- Embargos ?
- Quelles sont les licences qui explicitent les usages possibles
- Existe-t-il des procédures d'assurance qualité ?

Resources / sécurité

Garanties sur la faisabilité du plan

- Estimation des coûts
- Procédures de sauvegardes / récupération de données
- Mesures pour le long terme et les données sensible

Outils

Wizard for Data Management Plan Creation (experimental)



Documentation and Metadata

What information is necessary so that your research data is readable and interpretable in the future? Please describe how you document the primary research data. Here, the data should be described as specifically as possible in plain text. For measured data, please give for example the definitions of the variables; for annotated corpora, please include a glossary of all annotations as well as the annotation hand book.

Ex: The eye tracking data were captured using the model X from company Y with the following parameter settings. Participants in the experiment were presented with the following stimuli with the following instructions...

 [Tableau de bord](#) [Créer des plans](#) [DMPs publics](#) **Modèles de DMP** [Aide](#) [Langue](#) [emmanuelle Morlock](#)

CNRS

Modèles de DMP

Des modèles sont fournis par un financeur, un organisme, ou un tiers de confiance.

[Rechercher](#)

Nom du modèle	Nom de l'organisme	Type d'organisme	Description	Dernière mise à jour	Télécharger	Actions
CIRAD-TEMPLATE	CIRAD	Etablissement	Modèle pour réaliser un Plan de Gestion des Données. Version spécifique CIRAD basée sur le modèle H2020	20/06/2018	DOCX PDF	+ 📄
CIRAD-TEMPLATE-ENG	CIRAD	Etablissement	English version of the CIRAD model based on H2020 model for realizing a Data Management Plan.	25/10/2018	DOCX PDF	+ 📄
Horizon 2020 DMP	Commission européenne	Financier	<p>In Horizon 2020 a limited pilot action on open access to research data will be implemented. Projects participating in the Open Research Data Pilot will be required to develop a Data Management Plan (DMP), in which they will specify what data will be open. Other projects are invited to submit a Data Management Plan if relevant for their planned research.</p> <p>The DMP is not a fixed document; it evolves and gains more precision and substance during the lifespan of the project. The first version of the DMP is expected to be delivered within the first 6 months of the project. More elaborated versions of the DMP can be delivered at later stages of the project. The DMP would need to be updated at least by the mid-term and final review to fine-tune it to the data generated and the uses identified by the consortium since not all data or potential uses are clear from the start.</p> <p>The templates provided for each phase are based on the annexes provided in Guidelines on Data Management in Horizon 2020</p>	20/06/2018	DOCX PDF	+ 📄
Horizon 2020 FAIR DMP (anglais)	Commission européenne	Financier	<p>The Commission is running a flexible pilot under Horizon 2020 called the Open Research Data Pilot (ORD pilot).</p> <p>Projects participating in the pilot must submit a first version of the DMP (as a deliverable)</p>	31/10/2018	DOCX PDF	+ 📄

© 2016 - 2018 INIST • V2.0.5 • [A propos](#) [Conditions générales d'utilisation](#) [Tutoriels](#) • [Contact](#) [🌐](#) [🐦](#) [💬](#)

 **roadmap**

Catalogue de ressources

Non connecté Discussion Contributions Créer un compte Se connecter

Accueil Discussion Lire Voir le texte source Afficher l'historique Rechercher dans Cat OPIDoR

Cat OPIDoR, wiki des services dédiés aux données de la recherche

Quel type de service ? [modifier]

- INFORMATION
- FORMATION
- ACCOMPAGNEMENT
- OUTILS DE GESTION DES DONNÉES
- PLATEFORME D'ACQUISITION
- PLATEFORME DE CALCUL
- ENTREPÔT DE DONNÉES
- PLATEFORME D'ACCÈS
- PLATEFORME D'ARCHIVAGE

A quel stade du cycle de vie des données ? [modifier]

```
graph TD; Planification --> Collecte; Collecte --> Analyse; Analyse --> Documentation; Documentation --> Stockage; Stockage --> Conservation; Conservation --> Exposition; Exposition --> Réutilisation; Réutilisation --> Planification;
```

Dans quel domaine scientifique ? [modifier]

- SCIENCES HUMAINES & SOCIALES [Afficher]

Où ? [modifier]

Accueil

A propos

Modifications récentes

Naviguer par

- Type de service
- Stade du cycle de vie
- Domaine scientifique
- Service
- Structure d'appartenance

Contribuer

- Ajouter un service
- Ajouter une structure d'appartenance

Aide

- Description d'un service
- Description d'une structure
- FAQ
- Glossaire
- Cat OPIDoR en 2mn

Outils

- Pages liées
- Suivi des pages liées

[Home](#)[About](#)[Understand costs](#)[Compare costs](#)[Read more](#)[Vendor services](#)[Help](#)[Sign Up](#)[Sign In](#)

All about the costs of curation

What am I spending, what are they spending, what should we be spending?

Understand costs

Assessing your costs and using cost models to make smart investments



Compare costs

Add your curation costs and see how they compare with others

Grilles d'évaluation

<http://www.dcc.ac.uk/events/workshops/reviewing-data-management-plans>

Performance Criteria	Performance levels		
	Detailed	Addressed but incomplete / unsatisfactory	Not addressed
1. Summary of digital outputs and technologies	<p>Plan provides a clear description of the digital output or digital technology and the type of access envisaged e.g. 'freely available online'</p> <p>The summary explains how this will be achieved technically so reviewers can assess whether plans are appropriate.</p>	<p>There's a partial description for example an explanation of the data but not the access plans, or too little information given the scale and importance of the digital output. There is insufficient information to assess whether the overall plans are sound.</p>	<p>It is unclear what will be created or how.</p> <p>The proposal indicates that significant digital outputs will be created but the technical plan doesn't match this.</p>
2a. Standards and formats	<p>Relevant statistics are provided to explain the size, quantity and duration of the data.</p> <p>An explanation is given about which formats and standards will be used and why. This demonstrates an understanding about which formats are appropriate to each purpose e.g. to provide access or to archive data.</p>	<p>A data volume is stated, but it's difficult to interpret how many files that pertains to or how large and complex the dataset is.</p> <p>Formats and standards are named but it is not clear that these are understood and that choices have been made for valid reasons.</p>	<p>The description misses important information and doesn't give a clear sense of sounds decisions regarding the technical methodology.</p>
2b. Hardware and software	<p>It is clear what (if any) additional hardware and software will be required. The additional kit is clearly necessary and aligns with statements in the justification of resources.</p> <p>If nothing is required, the plan should clearly state this is not applicable.</p>	<p>The suggested resources seem unconvincing (e.g. items that should already provided by the institution or software that doesn't seem necessary for the proposed digital output).</p>	<p>The plan does not state that this section is not applicable.</p>
2c. Technical methodology	<p>There is a full description of the development process from data capture to delivery of a digital resource. Key issues such as quality control and backup are covered for all the research environments / circumstances e.g. in-house procedures and fieldwork.</p> <p>There is a clear timetable for technical delivery that relates to the overall project milestones in the case for support.</p>	<p>Certain information is missing e.g. how quality will be cheked if a number of assistants are transcribing and documenting resources.</p> <p>Backup is not explained for all circumstances in which data will be collected or stored.</p> <p>A timetable is missing, so it is unclear how the technical work aligns with the research activity.</p>	<p>There is no clear technical methodology or plan for how the data will be acquired, processed and used.</p>

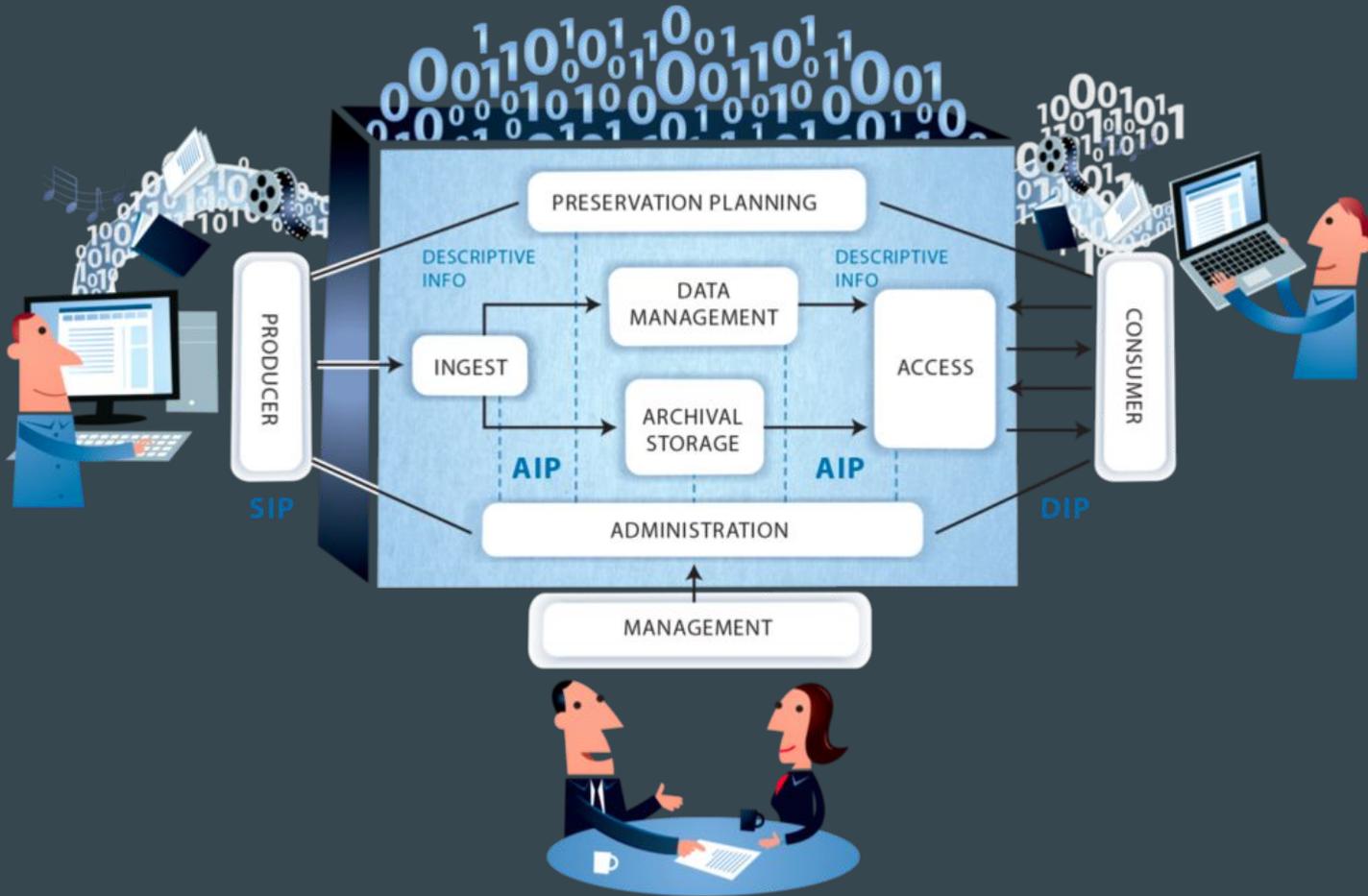
Bonnes pratiques pour contenir les coûts

Intégrer ces opérations scientifiques de création / traitement des données

- Recueil de consentement sur la diffusion
- Description (qualifier les variables, utiliser des standards)
- Nettoyage des données (procédures de contrôle qualité)
- Attribution de licences claires

L'archivage numérique à long terme ?

Ne pas confondre “stockage sécurisé” de données et archivage numérique à long terme (archivage dit “pérenne” : plus de 30 ans).



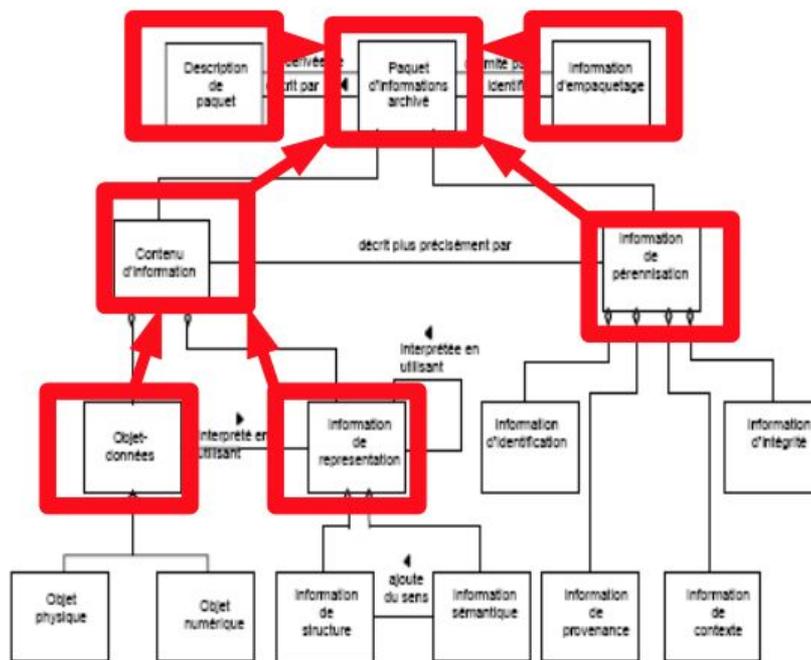


Schéma 4-18 : Paquet d'informations archivé (Vue détaillée)

L'archivage pérenne de données numériques non préparées dans cette perspective est impossible.

Se former

Les données de la recherche dans les thèses de doctorat

Livre blanc

[https://hal.archives-ouvertes.fr/
hal-01192930/](https://hal.archives-ouvertes.fr/hal-01192930/)

Demandez des
formations à vos
écoles doctorales !

DONNÉES DE LA RECHERCHE APPRENTISSAGE NUMÉRIQUE

DES RESSOURCES POUR ACCOMPAGNER LA COMMUNAUTÉ SCIENTIFIQUE DANS LA GESTION ET
LE PARTAGE DE LEURS DONNÉES

TROUVER UNE RESSOURCE : VIDEO, FICHE, QUIZ...



[VOIR TOUTES LES RESSOURCES](#)

VOUS SOUHAITEZ RÉUTILISER LES RESSOURCES ?

Toutes les ressources sont librement réutilisables. Vous pouvez les adapter à votre contexte et nous contacter pour des réalisations sur-mesure !

[EN SAVOIR PLUS](#)

DONNÉES DE LA RE APPRENTISSAGE NU

DES RESSOURCES POUR ACCOM
LE PARTAGE DE LEURS DONNÉES

TRouver UNE RESSOURCE : VIDEO, FICHE, QUIZ...

- CONSULTER PAR THEMATIQUES
- 🔍 TROUVER DANS LE CATALOGUE
- ↻ REUTILISER
- ? FAQ
- 📖 GLOSSAIRE
- 🛠 OUTILS - SERVICES
- 📚 BIBLIOGRAPHIE - WEBOGRAPHIE
- 🕒 CONSULTER MON HISTORIQUE

- 👍 ENJEUX & BENEFICES
- ⚖ ASPECTS JURIDIQUES & ETHIQUES
- ✓ PLAN DE GESTION DE DONNEES
- 📄 METADONNEES
- 🔍 IDENTIFIANTS PERENNES
- 📦 DEPOT & ENTREPOTS
- 🗄 STOCKAGE & ARCHIVAGE
- 📄 DATA PAPERS & DATA JOURNALS
- 👤 ACCES & VISUALISATION

DANS LA GESTION ET

RESSOURCES

VOUS SOUHAITEZ RÉUTILISER LES RESSOURCES ?

Toutes les ressources sont librement réutilisables. Vous pouvez les adapter à votre contexte et nous contacter pour des réalisations sur-mesure !

EN SAVOIR PLUS

<https://doranum.fr/#>

Conclusion

Récapitulatif

- **But** : s'assurer que les données seront partageables et réutilisables
- **Double fonction** :
 - Faciliter/optimiser la gestion des données
 - Consigner les décisions impactant dans la conduite de projet et fournir un document de référence
- **Atout** : facilite la collaboration entre métiers :
 - chercheurs,
 - professionnels de l'information
 - informaticiens
- **Temporalité** : utile avant, pendant et après le projet
- **Métaphore** : le scénario du film
- **Echelle** ? le projet
- **Difficulté** ? **oui** (prendre les décisions en amont) vs **non** (rédiger le plan)

Enjeux

- identifier les risques le plus tôt possible dans le projet
- Décrire de manière explicite la manière dont vous vous conformez à la politique de données de vos financeurs et de votre institution ainsi que le respect des règles éthiques
- Intégrer des opérations spécifiques qui vont rendre possible le partage et la réutilisation des données
- Mieux gérer les données (faciliter, rationaliser, optimiser les coûts)
- Vous aider à conserver l'utilité des données et le stockage sécurisé pour usages futurs
- Rassembler des informations souvent éparses de manière officielle : document de référence sur les données

Outil stratégique d'aide à la décision

Le PGD offre une vue d'ensemble, un tableau de bord pour :

- Identifier les **risques** (perte, qualité, confidentialité, coûts)
- Mettre en relation cycle de vie et **coûts**
- Anticiper **l'automatisation** des opérations d'intendance (conversions, transferts, dépôts)
- Faire ressortir les jeux de données à fort **impact**
- Faire des **choix** informés (points de vue chercheurs, prof. de l'information, informaticiens)
- Définir des **stratégies** de publication de données (dépôts et data papers)

Outil de collaboration

- Comme un scénario de film, permet à tous les personnels impliqués de partager une vision commune et d'apporter son expertise.
- Multiples acteurs / métiers
 - chercheurs titulaires
 - doctorants
 - post-docs
 - Ingénieurs soutiens
 - Documentalistes
 - Secrétaire d'édition / de rédaction
 - Archivistes
 - Informaticiens (projets, plateformes)
 - Ingénieurs projet (Europe)
 - ...

Le partage au premier plan des finalités

