



www.cnrs.fr

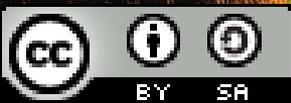
Participer à l'organisation du management des données de la recherche : gestion de contenu et documentation des données

Formats et métadonnées

ANF



Catherine Morel-Pair
Inist-CNRS, service Publier
catherine.morel@inist.fr





www.cnrs.fr

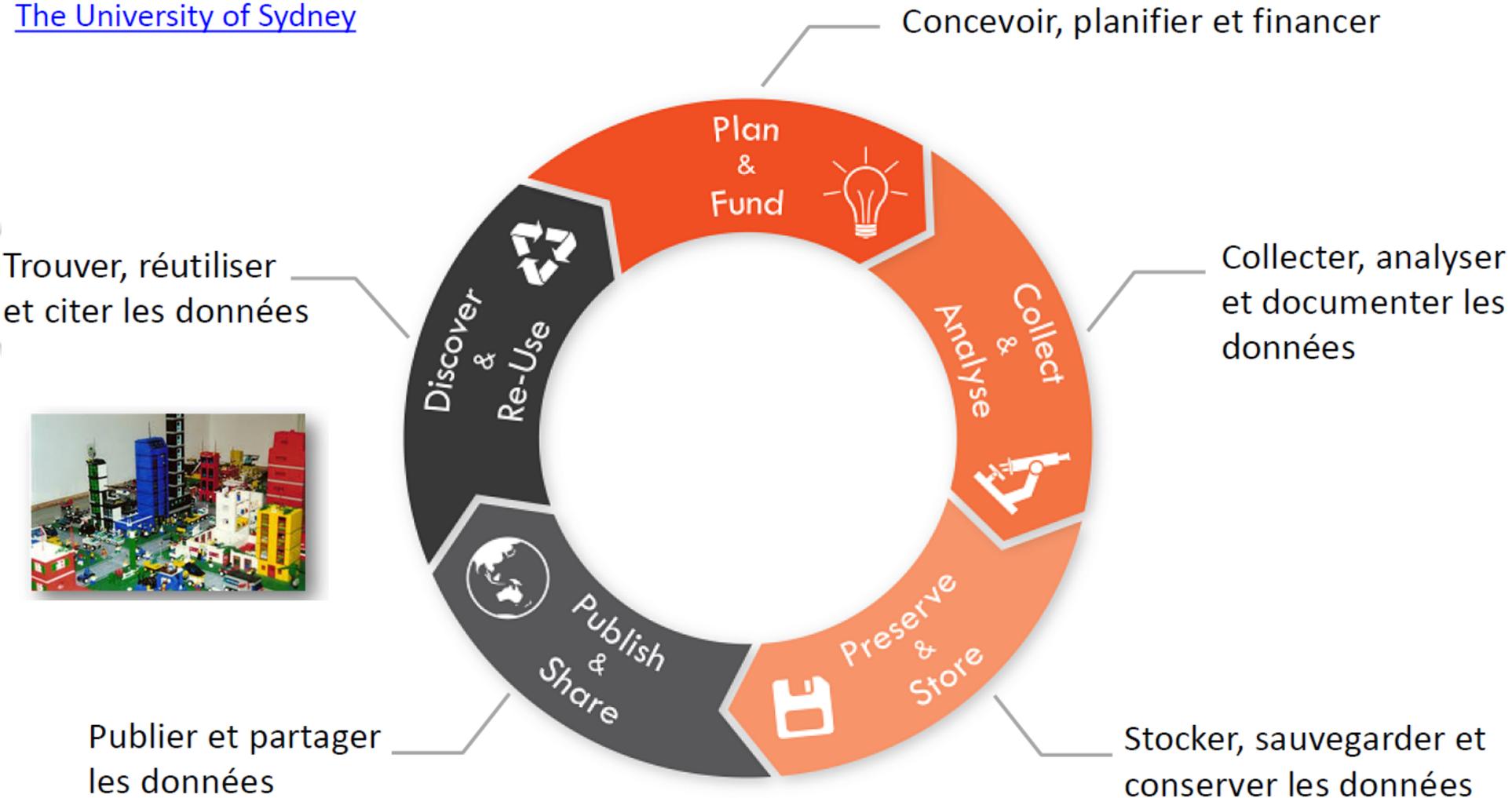
Introduction

- *Données de recherche : cycle de vie*
- *Des « FAIR data » pour l'Open Science*
- *Interopérabilité ?*
- *DMP : “formats” et métadonnées*

Données de recherche : cycle de vie

Source : (10), adapted from

[The University of Sydney](#)



Des “FAIR data” pour l’Open Science

- [Lignes directrices pour la gestion des données dans Horizon 2020, annexe 2](#)

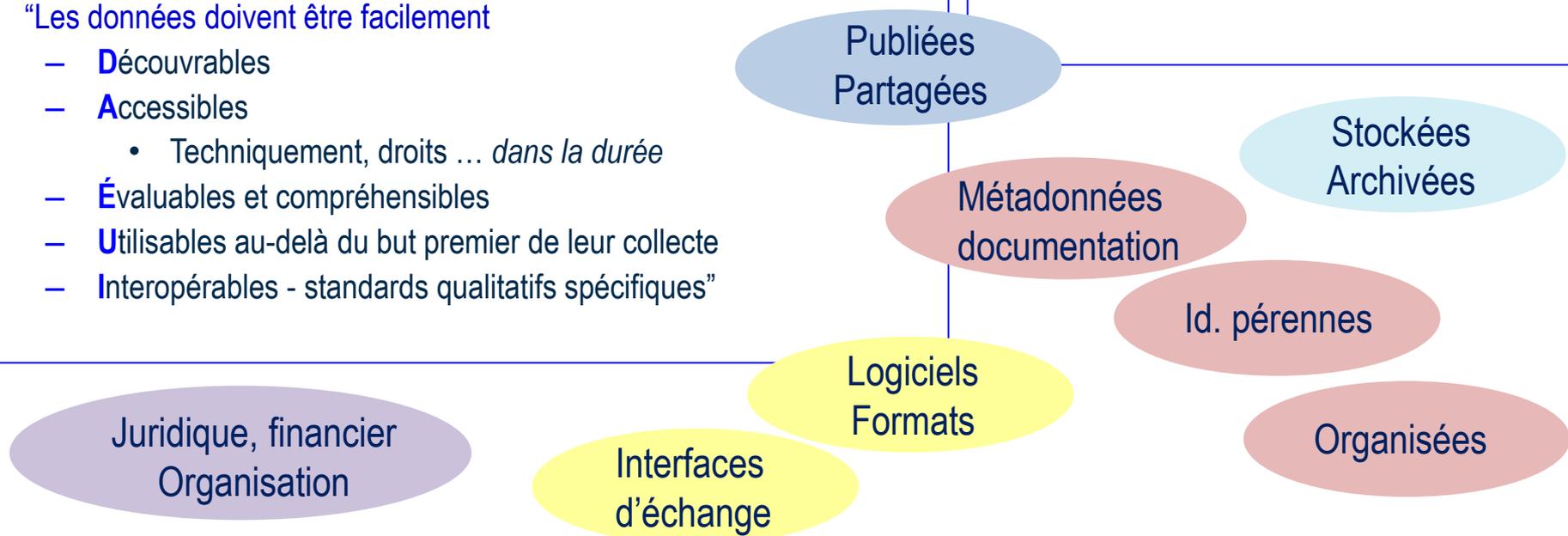
“ Les présentes prescriptions peuvent s’appliquer à n’importe quel projet qui produit, collecte ou traite des données de la recherche, et ces indications sont ajoutées au présent document comme support de référence lors de l’élaboration de plans de gestion des données dans le cadre de projets d’Horizon 2020.”

- “Les données doivent être facilement

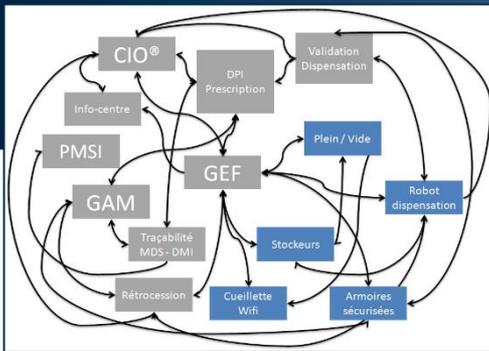
- **D**écouvrables
- **A**ccessibles
 - Techniquement, droits ... *dans la durée*
- **É**valuables et compréhensibles
- **U**tilisables au-delà du but premier de leur collecte
- **I**nteropérables - standards qualitatifs spécifiques”

- [Open Science, Amsterdam Call for Action](#)

- Data should be
 - **F**indable
 - **A**ccessible
 - **I**nteroperable
 - **R**e-usable.



Interopérabilité ?



Source : [Copilote, solutions santé](#)

Capacité d'une ressource ou application à fonctionner dans un autre système, un système plus vaste

- Capacité que possède un produit ou un système, dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs et ce sans restriction d'accès ou de mise en œuvre.

Notion critique dans de nombreux domaines d'activité

- Les différents systèmes, appareils et éléments utilisés, fabriqués par des constructeurs différents, doivent fonctionner sans heurt
- Implique des normes, standards, recommandations permettant la communication entre eux ([Wikipedia fr](#))
- France : [référentiel général d'interopérabilité](#) [informatique], RGI
- Publications et données scientifiques
 - Données, acteurs, outils, formats, projets ... multiples, divers, évolutifs
 - Un des objectifs des DMP / PGD et de H2020 → sections sur les **données**, leurs **métadonnées** et les **systèmes informatiques** qui les gèrent

DMP : formats et métadonnées

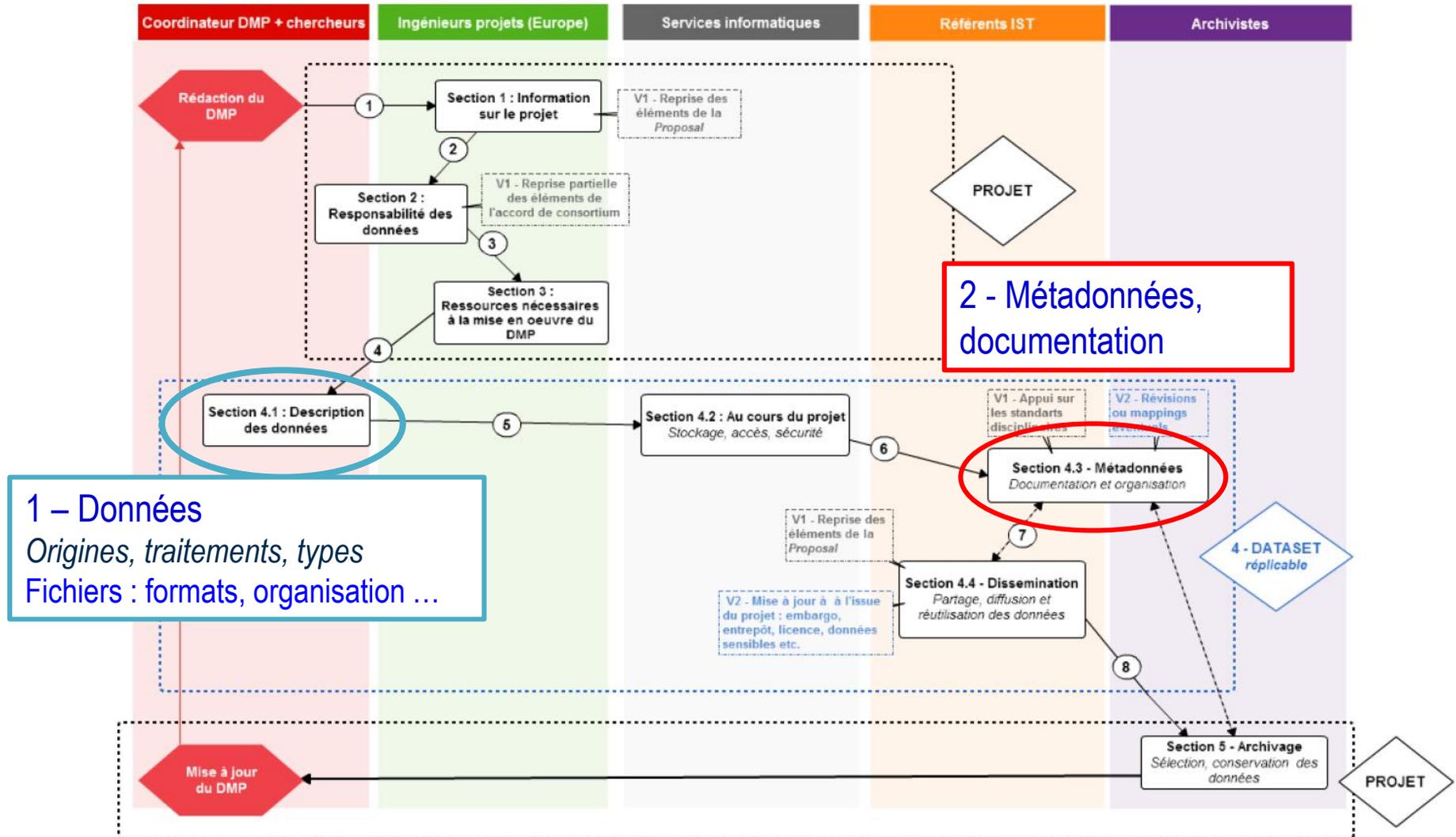


Schéma d'un DMP, exemple H2020 A.CARTIER, M.MOYSAN, N.REYMONET. [Réaliser un plan de gestion de données : guide de rédaction](#) (V1, 09/01/2015)

Sommaire

Partie 1 : Fichiers de données

- ① 1.1. Organisation et nommage
- ① 1.2. Formats

Partie 2 : Métadonnées et documentation

- ① 2.1. Introduction, définitions
- ① 2.2. Métadonnées : des standards
- ① 2.3 La documentation des données
- ① 2.4. Excursions sur sites ...
- ① 2.5. Le mot de la fin

Références

Des questions qui impliquent en complémentarité de compétences un travail d'équipe chercheurs, informaticiens et documentalistes





www.cnrs.fr

1 – Fichiers de données

1.1. Organisation et nommage

1.2. Formats

Les données numériques : des objets délicats

- Un jeu de données (dataset) numérique = de multiples fichiers
 - De types/contenus, formats informatiques et qualité variés
 - Evolutifs : accroissement, traitements, versions successives
 - Relations entre fichiers
 - Un **cycle de vie** ... Plusieurs acteurs
 - Sur support **informatique**, géré par des outils informatiques
 - Règles et limites, risques
- ... **Bonnes pratiques sur formats, organisation, nommage**
 - **Objectifs** : pouvoir utiliser et réutiliser les données dans tout leur cycle de vie
 - Dans un “dataset” : retrouver les fichiers pertinents – comprendre, distinguer
accéder au contenu, le manipuler
 - Pérenniser : stocker/sauvegarder de manière utile ... archiver ...



Credits : [jmason](#)

Organisation et nommage

...adaptés au projet et à ses acteurs

Une arborescence



Répertoires ou “Dossiers”

Sous-répertoires, sous-dossiers

Fichiers de données et de documentation, outils

cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputeeu_paraverbal111_dl.rtf	03/06/2015 13:56	Format RTF	38 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputeeu_paraverbal112_dl.rtf	03/06/2015 13:56	Format RTF	28 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputeeu_paraverbal113_dl.rtf	03/06/2015 13:56	Format RTF	33 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputefr_paraverbal01_dl.rtf	03/06/2015 13:56	Format RTF	81 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputefr_paraverbal02_dl.rtf	03/06/2015 13:56	Format RTF	124 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputefr_paraverbal03_dl.rtf	03/06/2015 13:56	Format RTF	97 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputefr_paraverbal04_dl.rtf	03/06/2015 13:56	Format RTF	205 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputefr_paraverbal05_dl.rtf	03/06/2015 13:56	Format RTF	63 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputefr_paraverbal06_dl.rtf	03/06/2015 13:56	Format RTF	82 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputefr_paraverbal07_dl.rtf	03/06/2015 13:56	Format RTF	35 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputefr_paraverbal08_dl.rtf	03/06/2015 13:56	Format RTF	34 Ko
cdsp_bequali_sp3_col_transcr_entretienindiv_fr_deputefr_paraverbal09_dl.rtf	03/06/2015 13:56	Format RTF	54 Ko

Enregistrer sous : Rapport Northwind-Final

Tags :

Emplacement : Office

Emplacements en ligne

Format du fichier : Document Word (.docx)

Nom : caracteres_signifiants_choisis.extensionFormat

L'arborescence

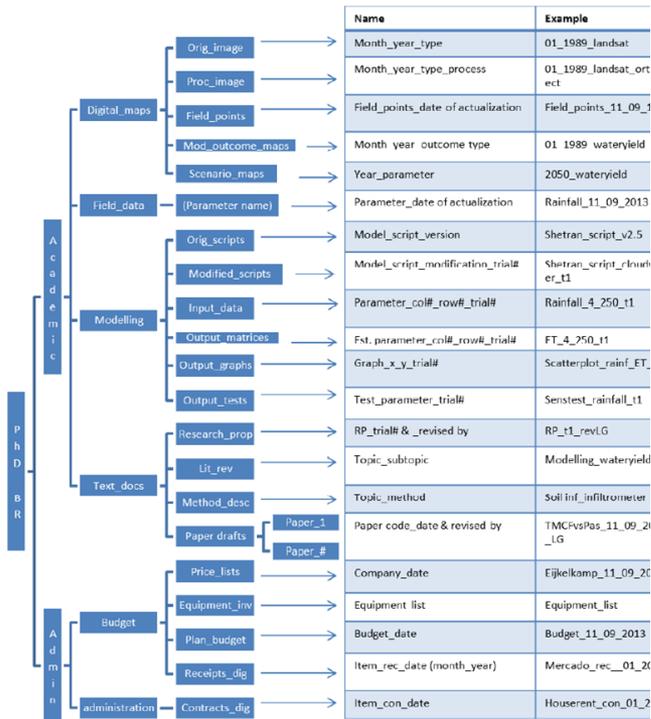


Figure 1. Proposed folder arrangement and file naming strategy.

Source : (1)

- Equivalent d'un plan de classement
 - Qui fait sens → orientée usages et consensuelle
 - Réflexion, modélisation préalable
 - 3 à 5 niveaux
- Dossier/sous-dossiers pour les aspects “support”
- Dossier(s) données / datasets
 - Type de données : textes, images, modèles
 - Période de temps
 - Caractéristiques du sujet d'étude : espèces, groupe d'âge ...
 - Activité de recherche : interview, étude, expérimentation
 - Brutes/traitées/dérivées
- Dossier pour la documentation scientifique et technique
 - Méthodologie, publications, rapports, DMP ...

Documenter l'arborescence et les fichiers ?

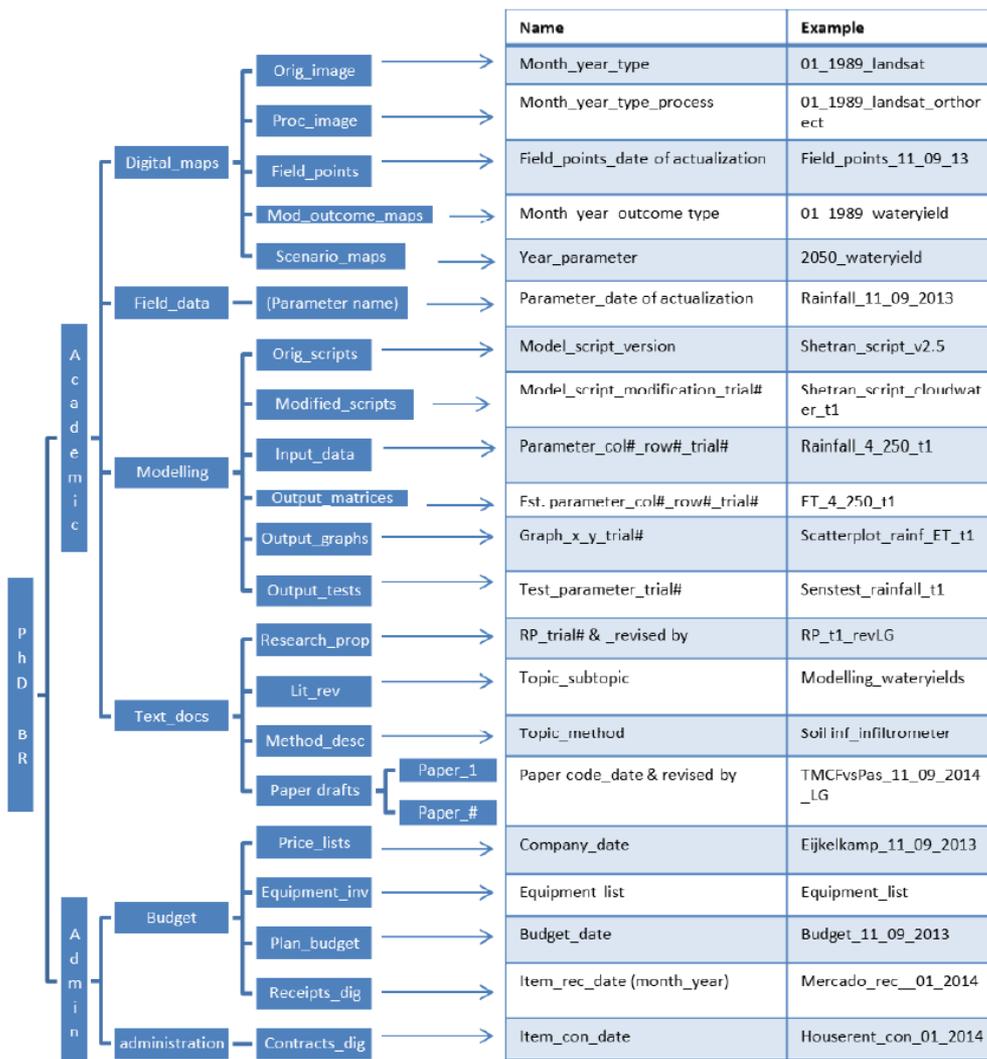


Figure 1. Proposed folder arrangement and file naming strategy.

Source : (1)

- Bonne pratique : **documentation technique sur les fichiers**, autant que nécessaire
 - Utile pendant le projet
 - Et pour réutiliser

“Figure 1 presents a diagram of folder organization and file naming strategy according to data categories (Table 1). In each folder, a metadata a text file (.txt) will contain relevant information of its files. For example, in the folder Orig_image, it is important to know for each image its acquisition date, satellite employed, source of the image and pre-processing procedures (if any). This diagram can also be used as a data index of the project, and can facilitate communication with other users interested in using/viewing this project's information.”



1 – Fichiers de données

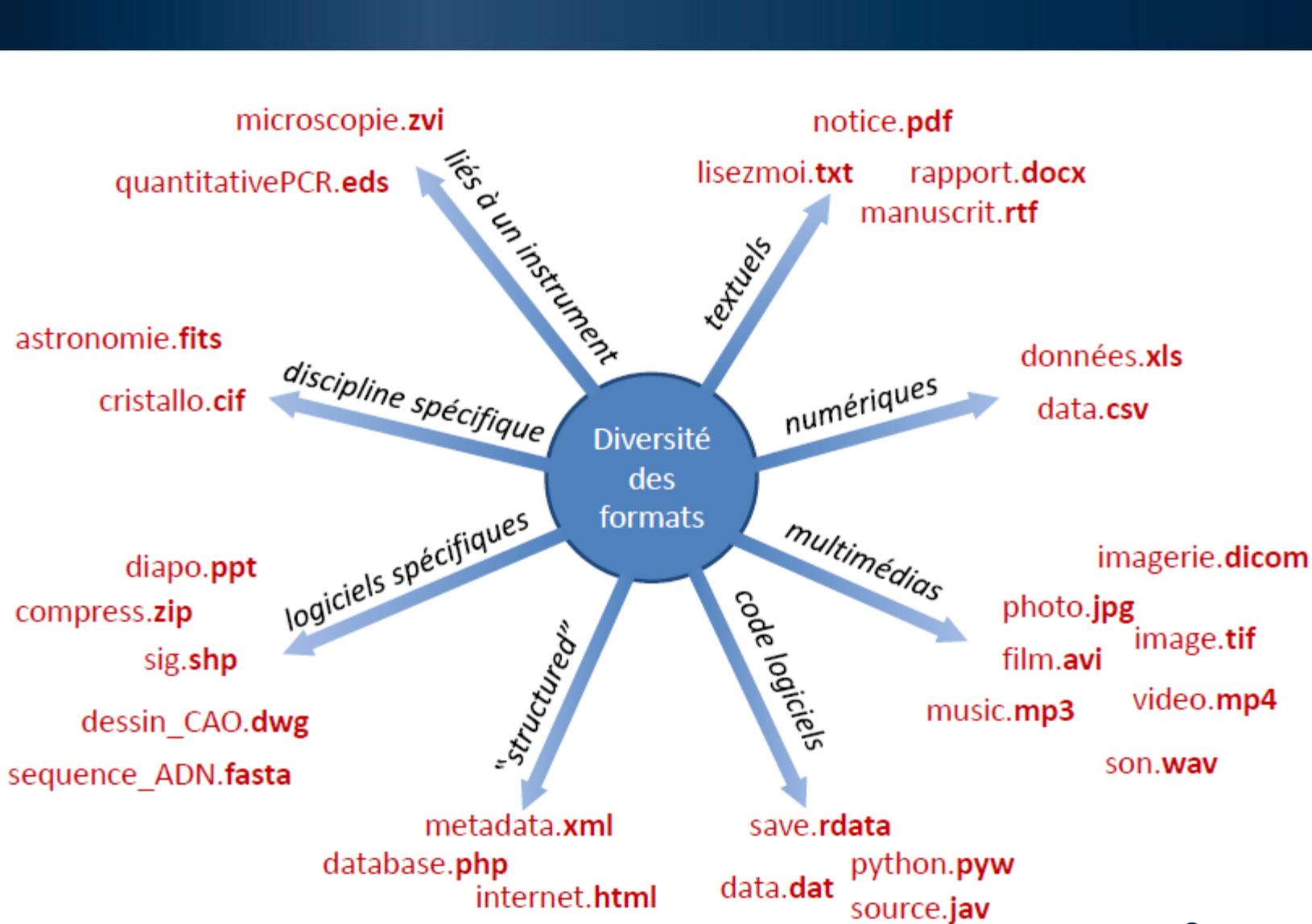
1.1. Organisation et nommage

1.2. Formats



www.cnrs.fr

Formats



bmp
cgm
emf
example
fits
g3fax
gif
ief
jp2
jpeg
jpm
jpx
ktx
naplps
png
prs.btif
prs.pti
pwg-raster
svg+xml
t38
tiff
tiff-fx
vnd.adobe.photoshop
vnd.airzip.accelerator.azv
vnd.cns.inf2
vnd.dece.graphic
vnd.djvu
vnd.dwg
vnd.dxf
vnd.dvb.subtitle
vnd.fastbidsheet
vnd.fpx
vnd.fst
vnd.fujixerox.edmics-mm
vnd.fujixerox.edmics-rlc
vnd.globalgraphics.pgb

Source : (10)

Formats, un challenge

- Un format informatique : une convention pour représenter une donnée sous forme numérique
 - Différents niveaux de convention/encodage entre le signal binaire et la donnée, voire l'information
 - Règles syntaxiques (structure) et sémantiques
 - Déterminant pour les négociations homme-machine et machine-machine : logiciels de représentation/traitement des données et protocoles d'échange
 - Format indiqué par une signature interne et représenté par le suffixe "extension" du nom de fichier
- De nombreux formats, plus ou moins pérennes, plus ou moins accessibles ...
Et des versions successives de chacun // versions logiciel
- Comment
 - Prévenir l'obsolescence des fichiers ?
 - Pour garantir des données réutilisables car durables ?

FREQUENCY
Definition : plasma frequency
Unit : k Hz

Type : REAL
Range : 100 .. 9675
Length : 80 bits
Format FORTRAN : F10.2

```
1981-09-30T14:22:18.426Z    3946.38    2500.82    -1.71    -8.42    -19.33    -80.00    -80.00  
1981-09-30T14:22:22.585Z    3964.01    2591.92    -1.76    -8.83    -19.86    -80.00    -80.00
```

```
348.511205851981-09-30T14:26:26.134Z3553.843160.87-1.35-15.79-2  
91149.9917.8568.8872.2516.44624.938.671205851981-09-30T14:26:34
```

```
001000000010000000100000001100  
000100000001000000011010100000  
110010010110100110011001100000
```

Typologie des formats

- Un format de fichier repose volontiers sur des formats de plus bas niveau
 - HTML 3.2 ← SGML ← texte ← UTF-8 (ou ...) ← binaire
- Formats spécifiés
 - “suffisamment décrits pour en développer une implémentation complète” → outils
- Formats normalisés
 - Spécifications adoptées/reconnues par un organisme de normalisation (AFNOR, ISO, IEEE etc.)
 - Résultent souvent de l'association de grands acteurs
- Formats propriétaires
 - Cadre d'utilisation contrôlable par une personne ou une entité juridique
 - Brevet, copyright, droits d'auteur
 - Souvent fermé : spécifications techniques non publiques
 - Exemple : fichiers gérés par la suite Microsoft Office

Typologie des formats - 2

- **Formats ouverts**

- “format de données interopérable dont les spécifications techniques sont publiques et sans restriction d'accès ni de mise en œuvre.”
 - Légalement exemptés de droits d'utilisation
- A la fois spécifiés et accessibles
- Peuvent faire l'objet de normes
- Exemple : formats Open Office et Libre Office – images jpeg, png
- Un format peut être à la fois “propriétaire” et “ouvert” : certaines versions de PDF

- **Formats « conteneurs »**

- Contiennent (et font fonctionner ensemble) différents types de données
- Spécifications libres ou non
 - Données contenues : formats propriétaires ou libres
- Matroska (XML, ouvert) contient un flux vidéo (ex. H. 264), un flux Audio (ex. AAC), éventuellement une piste de sous-titre
- [HDF5](#) : données scientifiques, gros volumes
 - XML ouvert - outillé - [users](#) (NASA ... CIRAD, IFREMER ...)



Credits : [Gifi.fr](#)

Formats : critères d'interopérabilité-pérennité

- **Caractéristiques du format**
 - Ouvert, idéalement normé
 - Accessible : largement diffusé - Nombreux logiciels pour l'exploiter
 - Outils permettant de faire migrer le format vers un autre format
 - Indépendance
 - Vis-à-vis des autres formats, des systèmes d'exploitation, économique
 - Matérielle : non lié à un périphérique ou un support de stockage spécifique
 - **... Qualité intrinsèque des fichiers**
 - Parfois endommagé par une copie, un support vieillissant ...
 - Parfois logiciel de création ou de conversion peu rigoureux
- **Outils et recommandations ...**
- **Entrepôts de données actuels** : souvent peu contraignants mais objectif de faire progresser la communauté, les données → standards
 - Parfois outils de validation, correction

Registres de formats

image

Available Formats

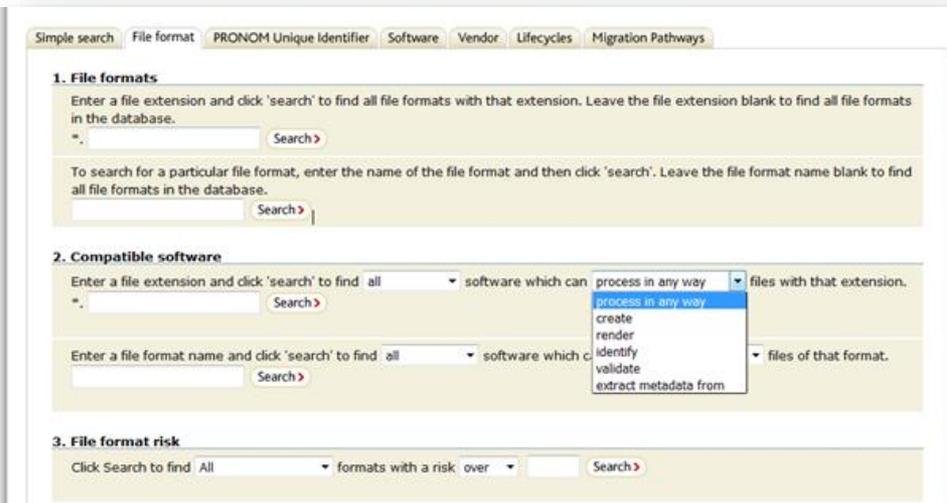


Name	Template	Description	Reference
bmp	image/bmp		[RFC-seantek-windows-image-03]
cgm	image/cgm	Computer Graphics Metafile	[Alan_Francis]
emf	image/emf		[RFC-seantek-windows-image-03]
example	image/example		[RFC4735]
fits	image/fits		[RFC4047]
g3fax	image/g3fax		[RFC1494]
gif			[RFC2045] [RFC2046]
ief		Image Exchange Format	[RFC1314]

- [MIME](#) ou [IMT](#)
 - IANA, autorité d'enregistrement du Web
 - Information brute : type → tableau, liens

- [The technical registry PRONOM](#)
 - Archive nationale du Royaume-Uni, actions “Digital préservation” et “Data management”
 - Base de données formats-software-éditeur-risque-dates-migrations ...
 - Recherche multicritères
 - Le plus complet + autres services

- Voir aussi : [List of file formats](#), Wikipedia
 - Orienté usages, catégorisation fine



Simple search | File format | PRONOM Unique Identifier | Software | Vendor | Lifecycles | Migration Pathways

1. File formats

Enter a file extension and click 'search' to find all file formats with that extension. Leave the file extension blank to find all file formats in the database.

Enter a file format name and click 'search' to find all file formats in the database.

2. Compatible software

Enter a file extension and click 'search' to find all software which can process in any way files with that extension.

Enter a file format name and click 'search' to find all software which can create render identify validate extract metadata from files of that format.

3. File format risk

Click Search to find All formats with a risk: over

Des orientations

Sous Catégorie	Standards
Caractère	UTF-8
Compression	Bzip2, gzip, ZIP, 7z, TAR
	ODF, OOXML, DocBook, PDF, PDF/A, EPUB3
	HTML, CSS, Internet media type, ATOM, APP, Javascript, CMIS
	XML, EXI XSD, JSON, OData, LDIF, RDF, OWL2, SPARQL, KML, DOM, SIARD, XMI, OAIS, SEDA
Description d'API	YAML, RAML
Identifiant	URI, ARK, ISNI
Géospatial	Shapefile, GeoJSON, GeoSpatial-Metadata, GML
Carnet d'adresse	vCard
Calendrier	iCalendar
	XSLT, XPath, XLink, XQuery, XInclude, XPointer, XML Signature
Géospatial	OpenLS, OWS Context, SLD
Conteneur vidéo	MPEG-TS, MP4, MKV, WebM
Codec vidéo	VP8, VP9, H.264, H.265
Conteneur audio	OGG
Codec audio	Opus, MP3, Vorbis, AAC, FLAC
Image	GeoTIFF, PNG, JPEG, SVG
	PAdES, XAdES, CAdES, ASIC
	IDMEF, IODEF

- Formats retenus ou conseillés dans un contexte d'usage - documentés, commentés
- [RGI 2015](#) , liste des formats retenus
 - Chacun est décrit plus loin + lien Wikipedia
 - + formats retirés, en fin de vie, en observation
- [Sustainability of Digital Formats, Planning for Library of Congress Collections](#)

**Sustainability of Digital Formats
Planning for Library of Congress Collections**

[Introduction](#) | [Sustainability Factors](#) | [Content Categories](#) | [Format Descriptions](#) | [Contact](#)

The Digital Formats Web site provides information about digital content formats. The analyses and resources presented here will increase and be updated over time. The compilers, Caroline R. Arms, Carl Fleischhauer, and Kate Murray invite [feedback](#) on the content.

Introduction
Background information and overview: What is a format? How shall we evaluate formats? What projects in other organizations are addressing these questions? >>

[Overview](#) | [Formats, Evaluation Factors, and Relationships](#) | [Papers and Presentations](#) | [Related Resources](#)

Sustainability Factors
What affects the ability of the Library to preserve content in a given format? These sustainability factors apply to all formats. >>

[Disclosure](#) | [Adoption](#) | [Transparency](#) | [Self-documentation](#) | [External Dependencies](#) | [Impact of Patents](#) | [Technical Protection Mechanisms](#)

Content Categories
The evaluation of formats must take into account quality and functionality. These factors vary according to the type of content under consideration and the categories will be expanded as time passes. >>

[Still Image](#) | [Sound](#) | [Textual](#) | [Moving Image](#) | [Web Archive](#) | [Datasets](#) | [Geospatial](#) | [Generic](#)

Format Descriptions
Documents with more information about specific formats. >>

[Browse categories](#) | [Browse alphabetical list](#) | [Format Descriptions as XML](#)

Des orientations - 2

CREATE & MANAGE DATA FORMATTING YOUR DATA

FILE FORMATS TABLE

The UK Data Archive works with different data formats for different purposes. There are optimal data formats that are used for long-term preservation of data.

This table contains guidance on file formats accepted by the UK Data Archive for deposited data. We welcome queries from researchers about appropriate file formats for working and preservation, particularly early in the research process.

Type of data	Acceptable formats for sharing, reuse and preservation	Other acceptable formats for data preservation
Quantitative tabular data with extensive metadata a dataset with variable labels, code labels, and defined missing values, in addition to the matrix of data	SPSS portable format (.por) delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) containing metadata information some structured text or mark-up file containing metadata information, e.g. DDI XML file	proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta) MS Access (.mdb/.accdb)
Quantitative tabular data with minimal metadata a matrix of data with or without column headings or variable names, but no other metadata or labelling	comma-separated values (CSV) file (.csv) tab-delimited file (.tab) including delimited text of given character set with SQL data definition statements where appropriate	delimited text of given character set - only characters not present in the data should be used as delimiters (.txt) widely-used formats, e.g. MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf) and OpenDocument Spreadsheet (.ods)
Geospatial data vector and raster data	ESRI Shapefile (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn) geo-referenced TIFF (.tif, .tiff) CAD data (.dwg) tabular GIS attribute data	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (KML) (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg)

- Create and Manage Data, [files formats table](#), UK Archives
- Entrepôts et projets de mutualisation, préservation ...
 - [Dryad](#), [formats recommandés](#) et [Biosharing](#), [standards](#)
 - CINES, avec ["Facile"](#) ...

Des outils d'identification, de validation ... de correction ... de conversion



Centre Informatique National de l'Enseignement Supérieur • Calcul intensif et Archivage pérenne

Accueil | Présentation | Calcul | **Archivage** | Services | Formations | Actualités | Annonces | Contact

Archivage > Des expertises > Expertise formats > FACILE

FACILE

Validation du Format d'Archivage du CINES par analyse et Expertise

Vérifier l'éligibilité de vos documents à un archivage sur la plateforme PAC du CINES, c'est **FACILE** !

Vous pouvez analyser leur degré de conformité à un format grâce à cet assistant en ligne, accessible librement, qui utilise le même processus de validation que la plateforme PAC.

Cet outil vous renseigne sur la liste des formats archivables et les outils utilisés pour les valider.

Il vous informe si un fichier est valide et bien formé par rapport au format que vous déclarez à l'aide d'une liste de choix.

Il vous suffit pour cela d'indiquer le fichier que vous souhaitez contrôler, puis de déclarer son format. Les outils de contrôle analysent votre fichier et renvoient automatiquement la réponse.

En cas de fichier mal formé ou non valide, vous pouvez directement faire appel à notre expertise par mail (analyse de second niveau) qui vous aidera à le corriger.

SOUS CETTE RUBRIQUE

Qu'est ce qu'un format de

Critères d'éligibilité d'un f

Analyse de fichiers avant

FACILE

Liste des formats archivab

ACTUALITÉS DE L

Le 24 juin 2011
services de
l'archivage

PREFORMA
PRESERVA
AVRIL 2011



FACILE - Service de validation de formats

Vérifier l'éligibilité de vos documents à un archivage sur la plateforme PAC du CINES.

Validation | Correction PDF | Tutoriels | Web Service

+ Choisissez des fichiers - Taille max 2,5 Go > Valider < Annuler

Cliquez ICI pour demander l'aide d'un expert du CINES

Détails	Fichier	Format identifié	Bien formé	Valide	Archivable dans PAC	Commentaire
	MS_1986_1_8.pdf	PDF 1.7	✗	✗	✗	Corriger automatiquement votre fichier Consulter nos tutoriels pour corriger votre fichier

Liste des formats validables

▲ Attention : le validateur de formats permet de valider certains formats qui ne sont pas pris en charge par la plateforme d'archivage du CINES.

Format	Nom	PRONOM PIUD	Type MIME	Commentaire	Archivable dans PAC
AAC AAC	Advanced Audio Codings	[fmt/199]		Format Mpeg-4 contenant uniquement un flux audio au format AAC.	✓
AIFF PCM	Audio Interchange File Format	[fmt/414]	[audio/x-aif, audio/x-aiff]	Format audio contenant uniquement un flux PCM.	✓
DAE UTF-8 1.4.1	Collada		[application/xml]	Format permettant de stocker des données géométriques sous forme de scènes (plusieurs objets combinés dans le même référentiel), et d'y ajouter des informations supplémentaires pour décrire la scène et les objets (matériaux, environnement lumineux, animations, ...) ou pour ajouter des notions sémantiques (relations entre les objets, découpage	✗

FACILE, CINES

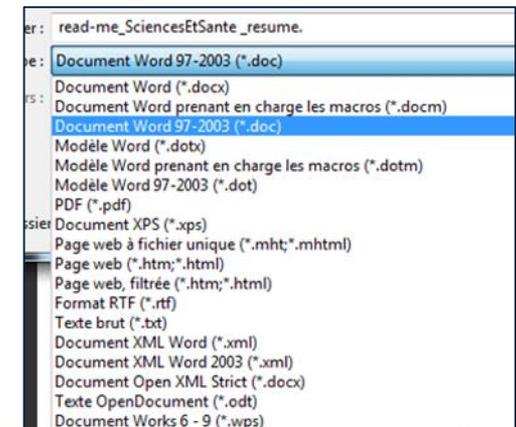
- Analyse les fichiers de formats archivables au CINES et plus
- Regroupe différents logiciels reconnus d'identification, validation – corrections
- “Chaînage des vérifications pour validation optimale”
- Interface Web & **Web service** - Documentation

Outils d'identification, validation, correction, conversion



Dernière modification le : 22 avril 2014

- Intégrés dans FACILE ... ou pas
- Annoncés | présents dans les projets disciplinaires, les grands projets d'infrastructure
- Accès web, installation en local pour certains : interface graphique et/ou ligne de commande → + de paramètres, lots de fichiers
- Outils de conversion
 - Voir aussi les formats exportés par les outils utilisés
 - Sites GitHub, SourceForge
 - ! Qualité du résultat → validation





2 – Métadonnées et documentation

2.1. Introduction, définitions

2.2. Métadonnées : des standards

2.3. La documentation des données

2.4. Excursion sur sites...

2.5. Le mot de la fin



www.cnrs.fr

Documenter : métadonnées et documentation

- Information permettant de trouver, comprendre gérer, vérifier, répliquer, réutiliser
 - “Qui, Quoi, Où, Quand, Comment et Pourquoi”

- Dans différentes étapes du cycle de vie
 - Pour soi et les autres
 - Information évolutive
 - ... Dès le début du projet



- Métadonnées
 - Données structurées → humains et machines
 - Des standards
- Documentation
 - Lisible par un humain. Des bonnes pratiques



2 – Métadonnées et documentation

2.1. Introduction, définitions

2.2. Métadonnées : des standards

2.3. La documentation des données

2.4. Excursion sur sites...

2.5. Le mot de la fin



www.cnrs.fr

Métadonnées

- Des éléments structurés
 - “nom d’élément” renseigné par une “valeur”
 - Exemples : Titre, Créateur, Format, Date ... + valeur
 - Destinés à des utilisateurs humains et à des machines
 - Outils de représentation, d’analyse, de conversion - Protocoles d’échange
 - Doivent être interopérables ... et pérennes
- ... Des standards, 3 niveaux
 - Vocabulaires standards d’éléments : Dublin Core, IPTC, EAD ... EML, ISO19115 ...
 - Valeurs de certains éléments : référentiels, identifiants, codes ISO ...
 - Syntaxes d’encodage : dans des tableaux (ex : csv), en “XML”, “RDF”, “JSON” ...
- ... Une typologie
 - Description du contenu “intellectuel”
 - Caractéristiques techniques des fichiers
 - Droits : producteur, d’accès, d’utilisations
 - Relations entre données, ex collections, versions



Métadonnées : standards et pratiques divers

- Objectifs, objets et acteurs

- Bibliothèque, documentation, archives
- NTIC, Web - Industrie et commerce
- ESR et données de recherche
- ... Big data and small data need metadata

- Complexité variable

- Des standards complémentaires en termes d'objectifs



Credits : culturevie.info

- Création

- Manuelle (saisie) ... Automatique (ex : métadonnées techniques) ... Par transformation d'un existant (extraction, conversions)

- Evolutives dans le cycle de vie

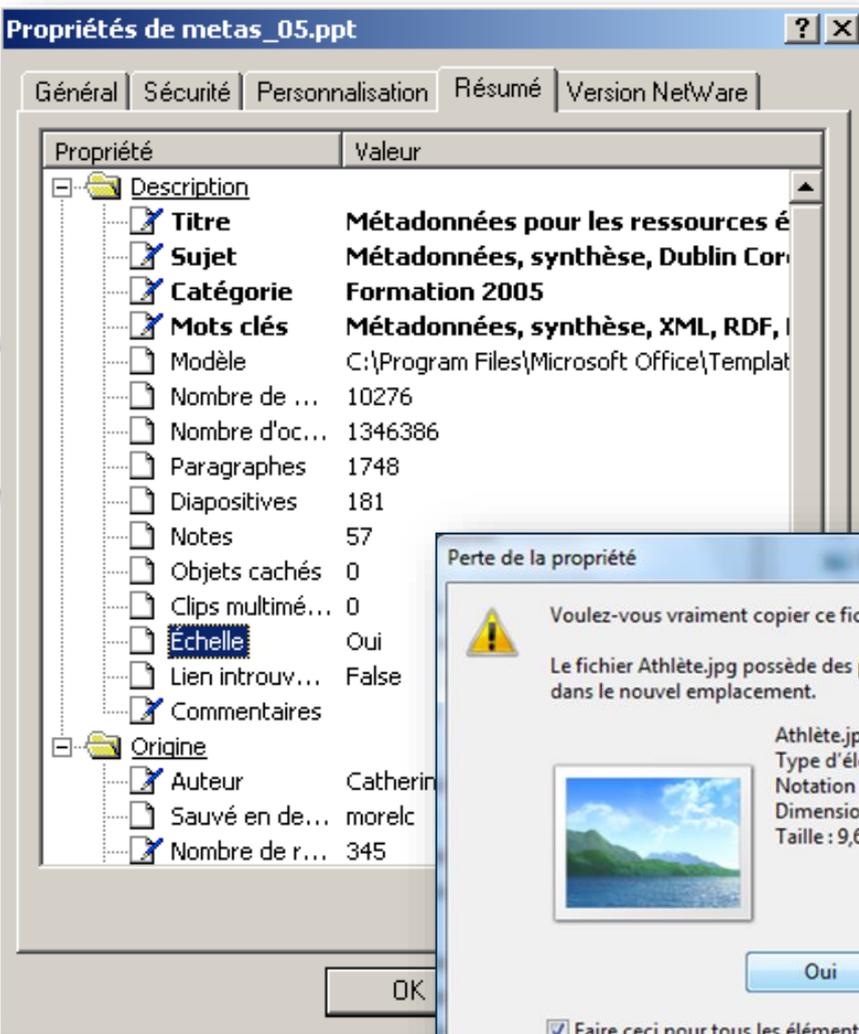
- Types des valeurs des éléments

- Texte, contrôlé ou modélisé / libre
- Autre : vignettes, représentation graphique ...

- Localisation

- Embarquées dans la ressource
 - Fréquent ... Utile, utilisé
- Externes : autre fichier
 - Indispensable pour rechercher, gérer, mutualiser

Des métadonnées courantes

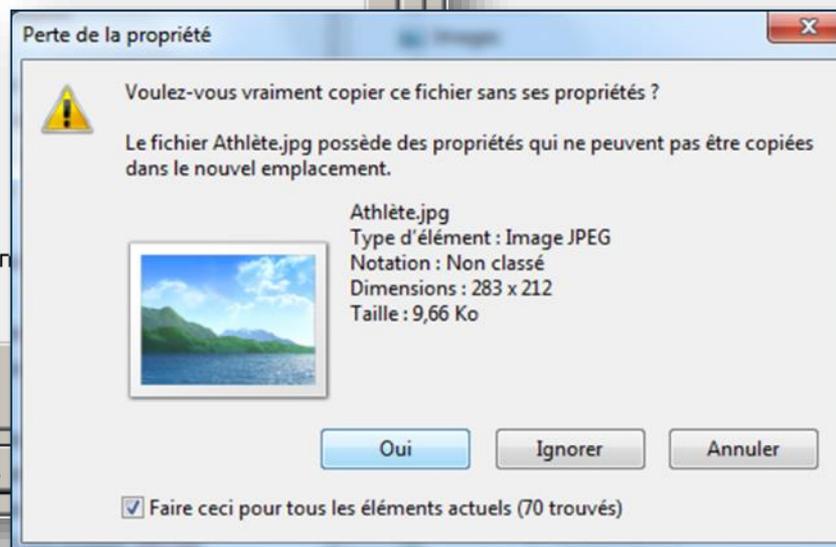


Métadonnées Windows

Explorateur Windows

Clic droit sur un fichier, « Propriétés »

... Dans le fichier (et des fichiers satellites)



! Variable entre versions

Catalogues

Signets
Liens sélectionnés par les bibliothèques universitaires

Rechercher un signet OK

SIGNETS

Flux rss tous les signets

SCIENCES ET TECHNIQUES

Sciences médicales (581)

Signets 1 à 10 / 581 | Page 1 / 59 | Aller à la page OK

Flux rss de cette thématique

 **Centre Français d'Evidence Based Dentistry (CFEBD)**
<http://cfebd.ups-tlse.fr/>

Basé à l'UFR d'odontologie de Toulouse, le CFEBD a pour objectif de fournir aux chirurgiens-dentistes, aux chercheurs et aux patients des informations récentes sur les données avérées de la science. Le CFEBD travaille en collaboration avec le groupe Cochrane pour la santé bucco-dentaire et traduit notamment les résumés des revues systématiques Cochrane.

Classification DEWEY : 617.6
Langues : français
Pays : France (métropolitaine)
Mots clefs : Odontologie, Stomatologie, EBM
Sélectionné par : Université Paul Sabatier - Toulouse 3
Propriété intellectuelle : UFR d'odontologie, Université Paul Sabatier (Toulouse 3)

[+1 +4](#) [INSIT.](#)
[Fiche LOM-FR](#)
[Ajouter un commentaire](#)

 **Agence régionale de santé**
<http://www.ars.midipyrenees.sante.fr/Internet.midipyrenees.0...>

Ouverte début avril 2010, l'Agence régionale de santé (ARS) de Midi-Pyrénées met en œuvre la politique régionale de santé, en coordination avec les partenaires et en tenant compte des spécificités de la région et de ses territoires.

Classification DEWEY : 614
Langues : français
Pays : France (métropolitaine)
Mots clefs : politique sanitaire santé publique
Sélectionné par : Université Paul Sabatier - Toulouse 3

[TOUS](#) [INSIT.](#)
[Fiche LOM-FR](#)
[Ajouter un commentaire](#)

 **Cancéro-pôle du grand Sud-ouest**
<http://www.canceropole-gso.org/>

Le cancéropôle contribue à coordonner la recherche en cancérologie au plus haut niveau d'excellence internationale, en particulier dans les domaines nouveaux issus de la révolution génomique et dans ceux des sciences sociales et économiques.

Classification DEWEY : 610 - 615.5 - 616.994

Catalogues de bibliothèques, d'entrepôts de données

Bases de données bibliographiques

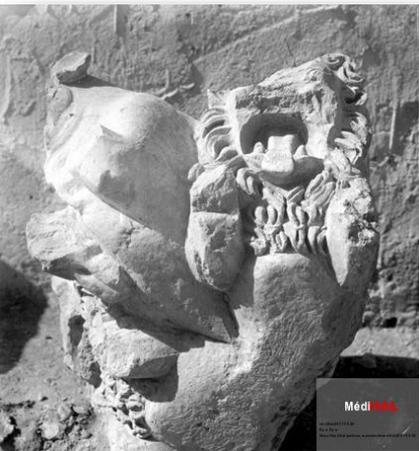
Portails, annuaires de sites

** Métadonnées "externes", localisation de la ressource

Métadonnées des images

Un cycle de vie, plusieurs représentations

1 - Métadonnées images “embarquées” dans le fichier image



Tag Name	Content
	---- IPTC ----
ObjectName	Fragments de deux chapiteaux à protomés trouvés dans le temple, sanctuaire de Bêl à Palmyre (Syrie)
DateCreated	1930
Headline	Fragments de deux chapiteaux à protomés trouvés dans le temple, sanctuaire de Bêl à Palmyre (Syrie)
By-line	Ifpo Ifpo
Keywords	Ifpo*Syrie*Palmyre*Tadmor*????*sculpture ornementale*calcaire de Palmyre*époque romaine*temple
Caption-Abstract	Mission archéologique française à Palmyre. Plaque (13x18)
Source	Ifpo-00266
City	Tadmur
Country-PrimaryLocationCo	SY
Country-PrimaryLocationNa	Syria
Credit	Ernest Will, Robert Amy ?

Content
---- EXIF ----
Canon
Canon EOS 300D DIGITAL
Horizontal (normal)
180
180
inches
2005:08:30 13:34:43
Centered
1/200
8.0
100
0221
2005:08:30 13:34:43
2005:08:30 13:34:43

Editables dans les logiciels spécifiques images :
Tableau “nom elt – valeur”

... Affichage brut : une syntaxe standard (XMP)

- Lisible dans le fichier par les machines équipées
- Exportable → autres usages

```
</rdf:Description>
<rdf:Description rdf:about="" xmlns:photoshop="http://ns.adobe.com/photoshop/1.0/">
  <photoshop:City>Tadmur</photoshop:City>
  <photoshop:Country>Syria</photoshop:Country>
  <photoshop:Credit>Ernest Will, Robert Amy ?</photoshop:Credit>
  <photoshop:DateCreated>1930</photoshop:DateCreated>
  <photoshop:Headline>Fragments de deux chapiteaux à protomés trouvés dans le
  temple, sanctuaire de Bêl à Palmyre (Syrie)</photoshop:Headline>
  <photoshop:Source>Ifpo-00266</photoshop:Source>
</rdf:Description>
```

Métadonnées des images - 2

2 – métadonnées **externes**, dans MediHal

medihal-00619945, version 1

PRÉVISUALISATION



1

LICENCE

Copyright (Tous droits réservés)

PARTAGER

f t+ 0

Fragments de deux chapiteaux à protomés trouvés dans le temple, sanctuaire de Bêl à Palmyre (Syrie)
1930
Ifpo Ifpo Détails

Description : Mission archéologique française à Palmyre. Plaque (13x18)

Domaine : Sciences de l'Homme et Société / Archéologie et Préhistoire

GÉOLOCALISATION



Données cartographiques ©2015 Basarsoft, Google, Maps GISrael, ORION-ME Conditions d'utilisation

Mots-clés : Ifpo Syne Palmyre Tadmor تدمر sculpture ornementale calcaire de Palmyre époque romaine temple

Format du dépôt	Fichier
Type de document	Image
Titre	fr Fragments de deux chapiteaux à protomés trouvés dans le temple, sanctuaire de Bêl à Palmyre (Syrie)
Auteur(s)	Ifpo Ifpo
Résumé	fr Mission archéologique française à Palmyre. Plaque (13x18)
Crédit	Ernest Will, Robert Amy ?
Pays	Syrie

Métadonnées affichées dans la page web - **externes** à l'image

- information plus accessible
- recherche, compréhension
- Enrichissements : contexte

MediHal, archive ouverte de collections d'images scientifiques
Dépôt :

- un formulaire de saisie de métadonnées
- un outil d'extraction de métadonnées standard

Source :

<https://medihal.archives-ouvertes.fr/medihal-00619945>

DMP : “Standards et métadonnées” ou “Métadonnées et documentation”

for example naming conventions, version control and folder structures. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.

Documentation and Metadata

What documentation and metadata will accompany the data?

Questions to consider:

- What information is needed for the data to be to be read and interpreted in the future?
- How will you capture / create this documentation and metadata?
- What metadata standards will you use and why?

Guidance:

Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed.

Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.

Ethics and Legal Compliance

How will you manage any ethical issues?

Questions to consider:

- Have you gained consent for data preservation and sharing?
- How will you protect the identity of participants if required? e.g. via anonymisation
- How will sensitive data be handled to ensure it is stored and transferred securely?

Guidance:

Ethical issues affect how you store data, who can see/use it and how long it is kept.

[Source : \(19\)](#)

- Quelles métadonnées seront utilisées pour accompagner les données
 - Standards de la discipline
 - Pourquoi ?
- Comment seront-elles créées/acquises ?
 - Partie automatisable ?

H2020 et les métadonnées des publications

Les bénéficiaires doivent aussi assurer un libre accès – via leur réservoir – aux métadonnées bibliographiques identifiant la publication déposée. Celles-ci doivent être dans un format standard et doivent faire figurer les éléments suivants :

- les termes [« Union européenne (UE) » et « Horizon 2020 »][« Euratom » et « programme Euratom de recherche et de formation 2014-2018 »];
- le nom de l'action, son acronyme et son numéro de subvention ;
- la date de parution et la durée de l'embargo (le cas échéant), ainsi qu'un identifiant pérenne.

... par exemple un DOI attribué par DataCite

Source: Lignes directrices H2020, [\(21\)](#), [\(22\)](#)

- Permet d'assurer visibilité du financement, suivi, production de statistiques, études d'impact
 - Fouille sur métadonnées plus efficace que sur texte intégral
- Et :
 - Date de parution et durée d'embargo
 - [Pas de cession complète des droits] - licence de réutilisation de type Creative Commons
 - Si possible, identification des contributeurs dans un système durable de type ORCID

H2020 et les métadonnées des données

(« données »), les bénéficiaires doivent :

- (a) déposer dans un entrepôt de données de la recherche et prendre des mesures afin de rendre possibles l'accès, l'exploration, l'exploitation, la reproduction et la diffusion par un tiers – et ce gratuitement pour tout utilisateur – des éléments suivants :*
 - (i) les données, y compris les métadonnées associées, nécessaires à la validation des résultats présentés dans des publications scientifiques le plus tôt possible ;*
 - (ii) les autres données, y compris les métadonnées associées, comme indiqué, et dans les délais stipulés dans le plan de gestion des données (voir annexe 1) ;*
- (b) fournir des informations – via l'entrepôt – sur les outils et instruments à la disposition des bénéficiaires et nécessaires à la validation des résultats (et, si possible, fournir les outils et instruments eux-mêmes).*

Source : [Lignes directrices ...](#) et [Lignes directrices pour le libre accès ...](#) H2020

- Et :
- Lien avec publications – “bien géré par Zenodo par exemple”
- Licence de type CC (CC0 ou CC-by) – “bien géré par B2SHARE” (infrastructure EUDAT)

Standards : trois niveaux

■ Sémantiques : vocabulaires de description standard

- DublinCore, MODS
- IPTC, EXIF
- EAD ...
- EML, ISA ...



■ Structure, syntaxes d'écriture

- Tableaux
- XML ... "RDF", JSON ...



• Exemples

- 200\$a[mon titre] vocabulaire MARC, format délimité
- <dc:title>mon titre</dc:title> vocab Dublin Core, format XML

- ... Bien gérés par les outils de publication numérique
 - Bases de données → formulaires de saisie ... imports et exports
 - ... Avec des règles du jeu à connaître ... +/-

■ Sémantique ++ : renseigner les éléments par des valeurs homogènes

- Référentiels matière, autorités ... Codes ISO ...
- Formes normalisées pour géolocalisation, date ...
- Valeur ajoutée fondamentale → interopérabilité

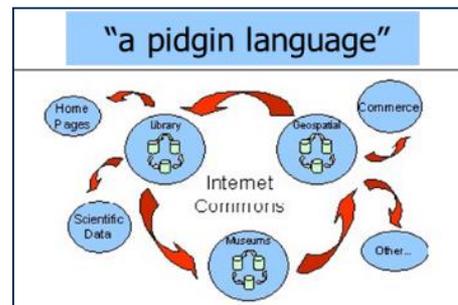




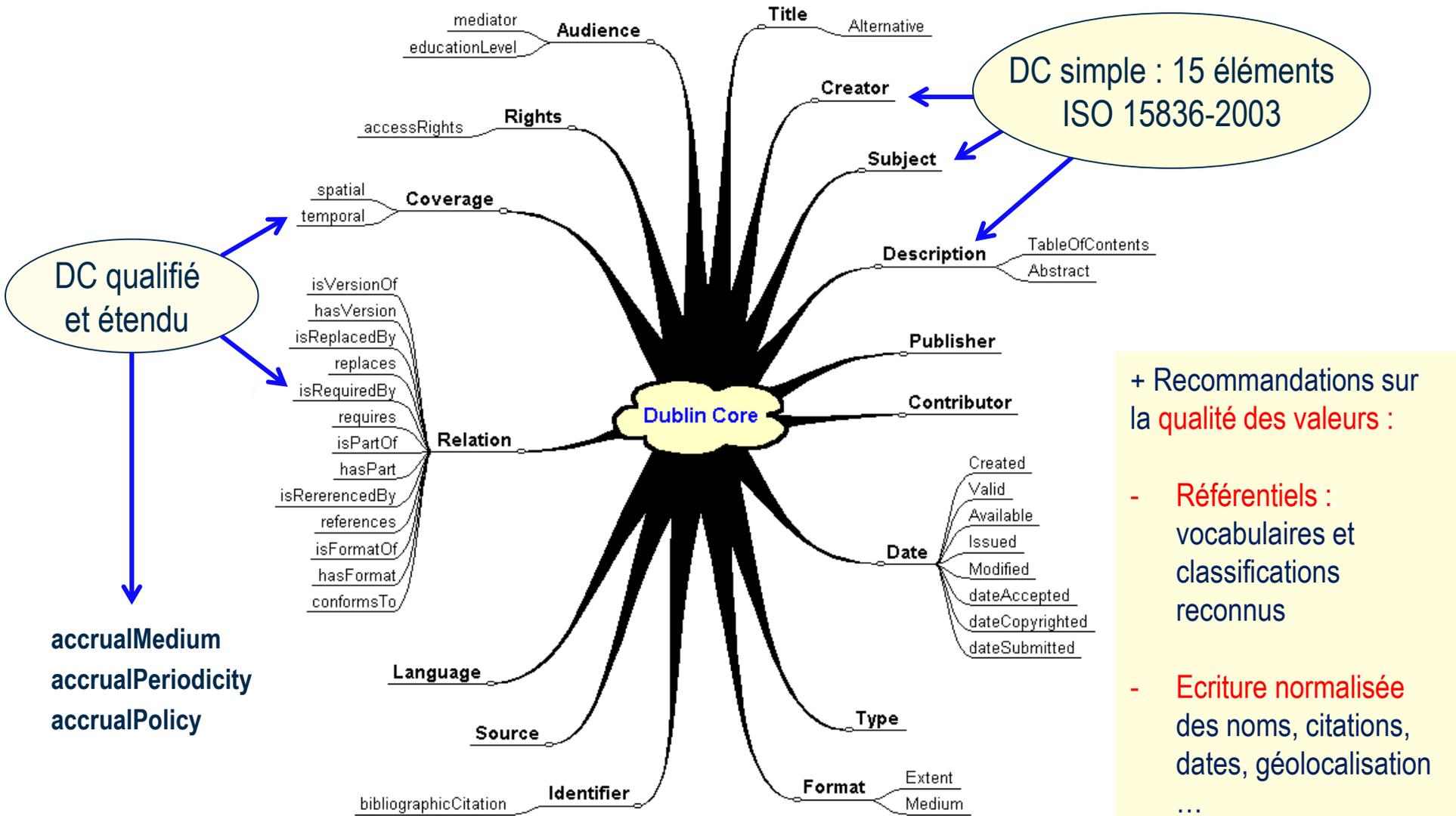
Un standard central : Dublin Core



- Dublin Core Metadata Initiative, 1995 <http://dublincore.org>
 - “Pallier les insuffisances des métadonnées HTML” pour donner une sémantique au Web [...] et améliorer la **découverte des ressources**
 - Par un **standard généraliste**
 - Une “description bibliographique **simplifiée**” pour tout type de ressource, de producteur
- Très utilisé pour mutualiser dans un monde ouvert
 - Consensus international et inter-professionnel
 - Standard de base de tous **les outils gérant des bibliothèques numériques**
 - “Ticket d’entrée” pour la mutualisation des archives ouvertes par **OAI-PMH**
 - Vocabulaire reconnu du “**Web de Données**” ...



Dublin Core



Source : <http://ganesha.fr/index.php?post/2008/03/31/Dublin-Core>

Documents en Français : [éléments](#), [qualificatifs](#)

Au-delà de Dublin Core...

- 1 – profils d'application de Dublin Core

- Ajouts

- D'extensions

- éléments supplémentaires issus d'un autre standard ou locaux,
 - attributs supplémentaires, précisant le sens d'un élément

- De contraintes : présence/nombre d'éléments, déclaration de référentiels particuliers pour les valeurs ...

- Adaptés aux besoins et usages ... et documentés

- Exemples

- Standard OLAC en linguistique

- Métadonnées de Dryad

- Métadonnées de DataCite

- DCAT (Data Catalog Vocabulary, W3C) et programme européen ISA d'infrastructure pour les catalogues de données et de services publics



Au-delà de Dublin Core... - 2

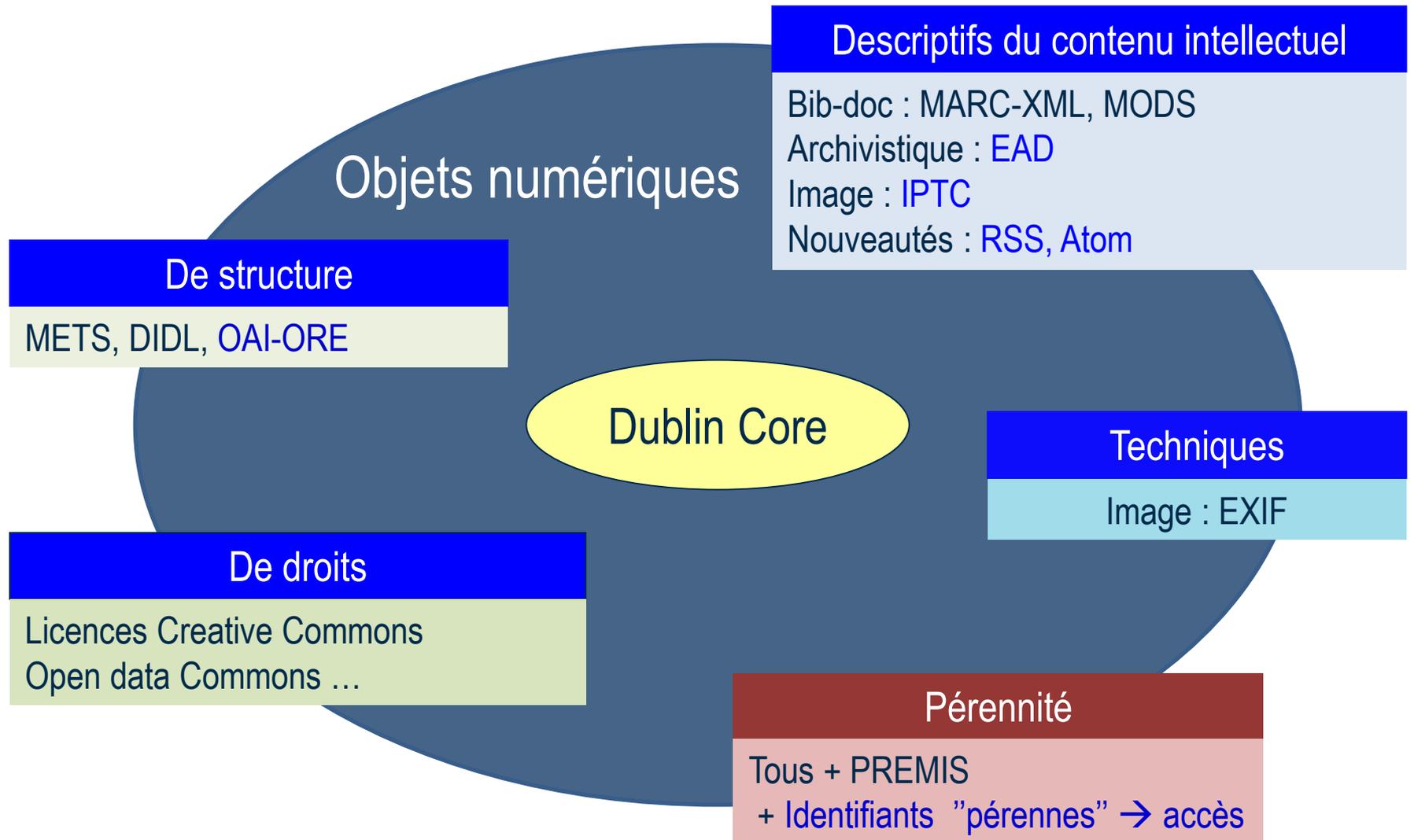
- 2 – des standards spécifiques

plus complexes, pour des contextes, informations et objectifs spécifiques

- Matériel d'étude et méthodes : processus d'acquisition et de traitement, qualité des données
- Evolutivité des contenus et des formats
- Organisation logique et physique des ensembles de fichiers, liens entre eux
- Contexte du projet, rôle des différents acteurs, questions éthiques
- Spécifications pour les réutilisations : outils associés, droits, questions éthiques ...
- ... Archivage pérenne, accompagné de bcp de métadonnées
 - Exemple : CINES, [Liste des métadonnées retenues pour le système d'archivage PAC](#), 2013



Quelques grands standards ... selon fonctions



Quelques grands standards ... selon les « objets » décrits

Objets numériques scientifiques

Types

Image : IPTC, EXIF
Multimédia : MPEG 4, 21
Métadonnées TEI ...

Disciplines

Linguistique : OLAC, IMDI, IsoCat
Etudes et enquêtes : DDI, SDMX
Géospatial : ISO 19115 et profils d'application
Génétique : ISA
Sciences évolution : Darwin Core
Ecologie : EML ...

Contributeurs

Référentiels et Id auteurs :
IdRef, VIAF, ORCID, ISNI
Contexte Web : vCards, FOAF ...

Valeurs des
métadonnées et
(méta)données
spécifiques

Référentiels "sujet"

Notices d'autorité classiques +
orientation web de données
Evénements : hCalendar,
Géoloc : geonames, KML ...
Sujets : thesaurus et
ontologies

Des valeurs d'éléments standardisées



- Une sémantique partagée → **interopérabilité** des catalogues, interprétation machines
- **Lier** les données, les acteurs
- **Référentiels sujets** : thésaurus, classifications, ontologies
- **Formes normalisées** langues, pays, dates, géolocalisation, formats ...
- **Identifiants pérennes** des données et des contributeurs
- **Référentiels matériel d'étude**, appareils, méthodes, types de données
- **Référentiels droits d'usage** : licences Creative Commons ...
- **Titre du dataset**
 - Métadonnée de recherche importante - trouvé → utilisé → cité - “What, Where, When, Who, and Scale”
 - Greater Yellowstone (**where**) Rivers (**what**) from 1:126,700 (**scale**) U.S. Forest Service (**who**) Visitor Maps (1961-1983) (**when**)
- **Citation**
 - Une option : le format standard de DataCite : **Creator**(**PublicationYear**): **Title**. **Publisher**. **Identifier**
 - Un **“formatteur”** **chez dataCite** → divers modèles de sortie

Des identifiants pérennes pour les données (PID)

URL

Mais un jour ...

The requested URL could not be retrieved

Des noms pérennes "actionnables"
sur le web par
des résolveurs de liens
Différents systèmes reconnus

ISSN, ISBN...

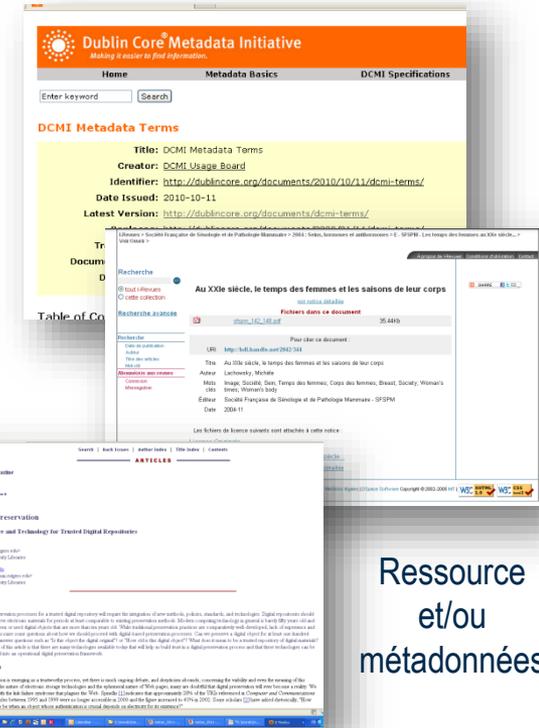
urn:nbn:de:gbv:7-isbn-90-6984-508-3-8
hdl:2042/344
doi:10.1045/june2005-jantz
ark:/12148/bpt6k3320625r

Résolveurs

<http://nbn-resolving.de/>
<http://hdl.handle.net>
<http://www.crossref.org/>
<http://gallica.bnf.fr/ark>

(+ résolveurs multi-systèmes)

Un critère d'interopérabilité



Ressource
et/ou
métadonnées

2016 : trois systèmes de PID très utilisés

Handle System®



Système ARK

- Monde libre scientifique - CNRI - très utilisé
- Délégation technique : serveurs locaux + qq serveurs centraux
- Tous types et formats, simple adhésion : 50 \$ par an
 - Isidore, Nakala, applications DSpace, infrastructure EUDAT ...
- DOI foundation et agences d'enregistrement
- Publications : [CrossRef](#), monde commercial édition puis plus
- [DataCite](#) : données de recherche ++ - un vrai projet
- BD centralisée, valeur ajoutée : qualité, recherche sur métadonnées, exports - Coûts variables – agence fr : Inist-CNRS
 - Zenodo, Dryad, KNB et ICPSR, etc ... – cité par H2020
- Univ Bib Californie et partenaires - libre
- Principes intéressants : pérennité, protocole, “granularité”
- Très décentralisé, peu outillé + un service central payant
 - En France : [BnF](#), Cines

+ *par disciplines ?*

- Comment choisir ? Avec décisions cycle de vie ... - Pas exclusifs l'un de l'autre
- Vraies questions : Granularité de l'identification - Dataset dynamiques ?

Identifiants et référentiels acteurs “Complémentarités et synergies”



- Bibliothèques et apparentés
 - Pratique ancienne : fichiers/référentiels d'autorité
 - [IdRef](#) (ABES) et [identifiants/référentiels BnF](#)
 - [VIAF](#), collaboratif : Bib nationales Europe puis plus
 - [IdHAL](#) et [AURéHAL](#)



- Identifiants chercheurs : [ORCID](#)
 - 2011, éditeurs privés + grands organismes publics
 - Auto-identification personnelle gratuite, conventions institutionnelles



- [ISNI](#)
 - Gestion des droits sur toute la chaîne de production-diffusion de biens culturels au sens large

■ Négociations/conventions VIAF-ORCID-ISNI ... et avec les institutions nationales

- Un Id entouré de méta-données chacun mais →
synergie : liens croisés, échanges de données

■ Des identifiants à utiliser et soutenir ...

■ Identifiants des éditeurs et des réseaux sociaux “recherche”

- ResearchGate, Academia.edu ...
- Stabilité et ouverture ?

Une syntaxe d'échange très utilisée : XML

- XML : format d'échange très répandu, 1998
- Ici, éléments Dublin Core en XML

```
<dc:title>La très haute nappe alluviale d'âge pléistocène inférieur de la vallée de la Creuse  
cryoturbation, habitats préhistoriques et datations absolues</dc:title>  
<dc:title>The very high alluvial formation of early Pleistocene age in the Creuse River V  
cryoturbation patterns, préhistorique occupation sites and absolute dating</dc:title>  
<dc:creator>Despriée, J.</dc:creator>  
<dc:creator>Gageaonnet, R.</dc:creator>  
<dc:subject>Vallée de la Creuse</dc:subject>  
<dc:subject>Formations alluviales</dc:subject>  
<dc:subject>Phénomènes de cryoturbation</dc:subject>  
<dc:subject>Tectonique récente</dc:subject>  
<dc:subject>Datations RPE</dc:subject>  
<dc:subject>Très ancien Paléolithique</dc:subject>  
<dc:subject>Plateau d'Aigurande</dc:subject>  
<dc:subject>Massif central</dc:subject>  
<dc:description>Malgré des conditions de prospections difficiles, de nombreux témoins  
de la Creuse ont été reconnus dans la région d'Eguzon, sur la bordure nord, très bocagère, du Massif
```

- Syntaxe ...
 - Sans importance pour déposer ou rechercher des données via formulaire de saisie ...
 - A connaître pour réaliser des imports par lots, utiliser des exports, résultats de requête API



2 – Métadonnées et documentation

2.1. Introduction, définitions

2.2. Métadonnées : des standards

2.3. La documentation des données

2.4. Excursion sur sites...

2.5. Le mot de la fin



www.cnrs.fr

“Documentation” ?

Source : entrepôt [Dryad](#)

DRYAD About ▾ For researchers ▾ For organizations ▾

Data from: Long-term data suggest jarrah-forest establishment at restored mine sites is resistant to climate variability



Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the [Dryad Terms of Service](#). To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.  [OPEN DATA](#)

Title	Standish et al. JEcol-2014-0252R1
Downloaded	5 times
Download	Standish et al. JEcol-2014-0252R1.csv (477.5 Kb)
Download	README.txt (4.64 Kb)
Details	View File Details

When using this data, please cite the original publication:

Standish RJ, Daws MI, Gove AD, Didham RK, Grigg AH, Kooh JM, Hobbs RJ (2015) Long-term data suggest jarrah-forest establishment at restored mine sites is resistant to climate variability. *Journal of Ecology*, 103(1): 78-89. <http://dx.doi.org/10.1111/1365-2745.12301>

Additionally, please cite the Dryad data package:

Standish RJ, Daws MI, Gove AD, Didham RK, Grigg AH, Kooh JM, Hobbs RJ (2015) Data from: Long-term data suggest jarrah-forest establishment at restored mine sites is resistant to climate variability. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.3m0rf>

[Cite](#) | [Share](#)

Files in this item

 Name: README.txt
Size: 4.64 Kb
Format: Text file
Description: README
Checksum (MD5): eb1233ec116d507f3db811f83031a968
[View/Open](#)

 Name: Standish et al. JEcol-2...R1.csv
Size: 477.5 Kb
Format: CSV file
Description: dataset file
Checksum (MD5): e4f68616cb260e397e3d81cd4536950
[View/Open](#)

- Information “moins structurée”
 - Sur le projet : hypothèse, méthodologie, échantillonnage, instruments, ...
 - Sur les fichiers ou bases de données et sur les “paramètres” : unités de mesure, vocabulaire, glossaire, abréviations ...
- Publiée avec les données
- Lisible par les humains : fichiers Readmefile.txt, document.pdf, .csv ...

Plus : [Managing and Sharing data, UK ARchives](#)



⊙ Eventuellement code informatique spécifique utilisé pour analyser les données

Documentation technique et projet - exemple : Bequali

Investaire de l'enquête « Quand des Français, des Anglais et des Belges (francophones) parlent d'Europe », A. Both © leQuali, 2013.

Questionnaires de sélection anonymisés pour les groupes anglais
 De spl_col_stat_fiche_gb_selection_questionnaire_001_d1.pdf 180 questionnaires
 à spl_col_stat_fiche_gb_selection_questionnaire_012_d1.pdf Numérotation continue en deux sections de 001 à 128 et de 200 à 253. Lacunes : _013, _127.
 De spl_col_stat_fiche_gb_selection_questionnaire_014_d1.pdf Datation estimée à mars 2006.
 à spl_col_stat_fiche_gb_selection_questionnaire_126_d1.pdf
 spl_col_stat_fiche_gb_selection_questionnaire_128_d1.pdf
 De spl_col_stat_fiche_gb_selection_questionnaire_200_d1.pdf
 à spl_col_stat_fiche_gb_selection_questionnaire_253_d1.pdf

Questionnaires complémentaires anonymisés pour les groupes belges
 De spl_col_stat_fiche_be_compl_questionnaire_001_d1.pdf 49 questionnaires
 De spl_col_stat_fiche_be_compl_questionnaire_011_d1.pdf Lacunes : _012, _042, _046, _53
 De spl_col_stat_fiche_be_compl_questionnaire_013_d1.pdf (fichiers non remplis, candidats non retenus). À noter que le _17 est incomplet (manque la dernière page) et le _49 non rempli (défection du participant).
 De spl_col_stat_fiche_be_compl_questionnaire_041_d1.pdf
 De spl_col_stat_fiche_be_compl_questionnaire_043_d1.pdf
 De spl_col_stat_fiche_be_compl_questionnaire_045_d1.pdf
 De spl_col_stat_fiche_be_compl_questionnaire_047_d1.pdf
 à spl_col_stat_fiche_be_compl_questionnaire_052_d1.pdf Datation estimée à février 2006.

Questionnaires complémentaires anonymisés pour les groupes français
 De spl_col_stat_fiche_fr_compl_questionnaire_005_d1.pdf 48 questionnaires.

INVENTAIRE DE L'ENQUETE "Le choix de l'école" par Agnès van Zanten (dir.) (code bequali si)

Date de mise à jour : 20/08/2015

beQuali

Numero	Identifiant de l'article / Nom de fichier	Fonds	Sous-fonds	Oes	Les	doces	Langue	Sujet	Titic (air le)	n°	Titre de l'article	Date	Niveau t/
0019	odsp_bequali_si_prop_methode_entre1_FR_0012_guide	odsp_bequali_si	prop	METHOS	0	0012	FR	0012	guide		Guide d'entretien thématique n°2 à destination des représentants d'associations de parents d'élèves, 2000-01-01	2001	rapport
0020	odsp_bequali_si_prop_methode_entre1_FR_0013_guide	odsp_bequali_si	prop	METHOS	0	0013	FR	0013	guide		Guide d'entretien thématique n°3 à destination des coordinateurs SEP, 2000-01-01	2001	rapport
0021	odsp_bequali_si_prop_methode_entre1_FR_0014_guide	odsp_bequali_si	prop	METHOS	0	0014	FR	0014	guide		Guide d'entretien thématique n°4 à destination des inspecteurs d'Éducation nationale, 2000-01-01	2001	MIS
0022	odsp_bequali_si_prop_methode_entre1_FR_0015_guide	odsp_bequali_si	prop	METHOS	0	0015	FR	0015	guide		Guide d'entretien thématique n°5 à destination des directeurs d'établissements scolaires, 2000-01-01	2001	rapport
0023	odsp_bequali_si_prop_methode_entre1_FR_0016_guide	odsp_bequali_si	prop	METHOS	0	0016	FR	0016	guide		Guide d'entretien thématique n°6 à destination des enseignants, 2000-01-01	2001	MIS
0024	odsp_bequali_si_col_transcr_entretien_FR_pirswil_poverne1	odsp_bequali_si	col	TRANSCR	1	0024	FR	0024	entretien	1	Transcription entretien n°1 de parents d'élève et notes de synthèse complémentaires, enquête Jules-Verne (Bequali), enquête des villes	2000-02-02	MIS
0025	odsp_bequali_si_col_transcr_entretien_FR_pirswil_poverne2	odsp_bequali_si	col	TRANSCR	1	0025	FR	0025	entretien	2	Transcription entretien n°2 de parents d'élève et notes de synthèse complémentaires, enquête Jules-Verne (Bequali), enquête des villes	2000-02-02	MIS
0026	odsp_bequali_si_col_transcr_entretien_FR_pirswil_poverne3	odsp_bequali_si	col	TRANSCR	1	0026	FR	0026	entretien	3	Transcription entretien n°3 de parents d'élève et notes de synthèse complémentaires, enquête Jules-Verne (Bequali), enquête des villes	2000-02-02	MIS
0027	odsp_bequali_si_col_transcr_entretien_FR_pirswil_poverne4	odsp_bequali_si	col	TRANSCR	1	0027	FR	0027	entretien	4	Transcription entretien n°4 de parents d'élève et notes de synthèse complémentaires, enquête Jules-Verne (Bequali), enquête des villes	2000-02-02	MIS
0028	odsp_bequali_si_col_transcr_entretien_FR_pirswil_poverne5	odsp_bequali_si	col	TRANSCR	1	0028	FR	0028	entretien	5	Transcription entretien n°5 de parents d'élève et notes de synthèse complémentaires, enquête Jules-Verne (Bequali), enquête des villes	2000-02-02	MIS

- Des inventaires documentés
 - utiles au cours du projet : échanges avec prestataires de numérisation, contrôles, suivi des transcriptions ...
 - Organisation, nommage, contenu des données publiées
 - Des **métadonnées dérivées par conversion et calcul**
- Un document projet complet “l'enquête sur l'enquête”
 - HTML, texte et vidéo et PDF

Analyse

méthodologie et résultats

Combiner interprétation et traitement systématique des entretiens



Les enquêtés mis à l'épreuve de la politisation



Des épreuves de politisation différemment ressenties



L'analyse principale a porté sur les 12 groupes de la première « famille ».

Plusieurs approches ont été mises en œuvre : des opérations interprétatives simples, une analyse lexicométrique et une analyse par codification à l'aide d'un logiciel de type CAQDAS.

Différents types de travaux ont été mobilisés : les travaux sur les rapports profanes et spécialisés au politique, sur la délibération, sur l'identité européenne et sur les attitudes à l'égard de l'UE (notamment les travaux par entretiens de Juan Diez-Medrano).

A été mise au centre de l'analyse l'idée que les enquêtés ont été confrontés à une épreuve de politisation, dans un cadre construit et semi-public. Ont été spécifiquement étudiés les manières de gérer le risque du conflit et les positionnements par rapport à un certain nombre de clivages.

>Le modèle de la politisation a été amendé

Extrait

Pour en savoir plus

In this section

Briefing Papers

How-to Guides

Developing RDM Services

Curation Lifecycle Model

Curation Reference Manual

Policy and legal

Data Management Plans

Tools

Case studies

Repository audit and assessment

Standards

Disciplinary Metadata

DIFFUSE

Publications and presentations

Roles

Curation journals

Informatics research

External resources

Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

Search by Discipline



Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

Search by Resource Type

Metadata Standards

Specifications for the minimum information that should be collected about research data in order to be re-used.

Profiles and Extensions

Standards that have been adapted for use in particular types of repositories, or for particular types of data.

Use cases

Institutional repositories and data portals using standards to determine which metadata should be collected for data deposit.

Tools

Software that has been developed to capture or store metadata conforming to a specific standard.



Standards de métadonnées, extensions, outils, études de cas :

DCC, [Digital Curation Center, Metadata Standards](#) par discipline et [General research metadata RDA, metadata Directory](#) avec dépôt [GitHub](#) des sources

Metadata

RDA | Metadata Directory

Edit this page

View the standards

View the extensions

View the tools

View the use cases

Browse by subject areas

Contribute

Add standards

Add extensions

Add tools

Add use cases

github

@twitter

linkedin

facebook

Life Sciences

[ABCD - Access to Biological Collection Data](#) [Edit](#)

The [Access to Biological Collections Data \(ABCD\) Schema](#) is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated.

Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.

[Darwin Core](#) [Edit](#)

A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.

Sponsored by Biodiversity Information Standards (TDWG), the current standard was last modified in October 2009.

[EML - Ecological Metadata Language](#) [Edit](#)

Ecological Metadata Language (EML) is a metadata specification particularly developed for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications).

Sponsored by ecoinformatics.org, EML Version 2.1.1 was released in 2011.

[Genome Metadata](#) [Edit](#)

Genome metadata on PATRIC consists of 61 different metadata fields, called attributes, which are organized into the following seven broad categories: Organism Info, Isolate Info, Host Info, Sequence Info, Phenotype Info, Project Info, and Others.

[ISA-Tab](#) [Edit](#)

The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies.

Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.

[MIBBI - Minimum Information for Biological and Biomedical Investigations](#) [Edit](#)

A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI

Pour en savoir plus - 2

HN Huma-Num
la TGIR des humanités numériques

Présentation ▾ Services et outils ▾ Consortiums International Ressources

À propos de
Huma-Num est un numérique de la re
Pour remplir cette à mettre en œuvre numériques pérenn de partenaires et d

- Présentation
- Stocker
- Traiter
- Exposer
- Signaler
- Diffuser
- Archiver

le recherche (TGIR) visant à faciliter le tournant et sociales.

est bâtie sur une organisation originale consistant certation collective) et technologique (services opéenne en s'appuyant sur un important réseau

Lire la suite

LES SERVICES D'HUMA-NUM

SIGNALER

Une plateforme d'accès et de référencement pour les données de la recherche en SHS

La TGIR Huma-Num développe un dispositif technologique qui permet d'accompagner les différentes étapes du cycle de vie des données numériques.

SUR NOTRE BLOG

22 juin 2015
ISIDORE Suggestions : il est temps de taguer dans vos billets !

1 juin 2015
De nouveaux entrants européens dans ISIDORE !

11 mai 2015
Avec Isidore, ajoutez des suggestions de lectures dans votre blog.

Plus sur le blog >>

SUIVEZ-NOUS SUR TWITTER

Stéphane Pouyllau @spouyllau 28 Jul
Qu'est-ce qu'une archive de chercheur ? | #Consortium Archipolis rechercheisidore.fr/search/ressour... + recension sur archipolis.hypotheses.org/224
Retweeted by Huma-Num
Expand

Huma-Num @Huma_Num 28 Jul
Découvrez la ferme d'outils de traitement pour les données numériques #SHS de @huma_num sur huma-num.fr/services-et-ou... #ecosystème
Expand

Consortium3D_SHS @Consortium_3D 24 Jul
Partenaire du #Consortium3D l'équipe EPOTEC (@IRCCyN UMR CNRS 6597 & #CFV (EA 1161)

Initiatives nationales, européennes, par discipline

[EUDAT](#), [“training”](#)

Research Data Alliance, [Recommandations & outputs](#)

[OpenAire guidelines](#)

[Huma-Num](#)
[“Services”](#) et [“ressources”](#)

[Biosharing, standards](#) ...

Equipe Données de la recherche, Inist



2 – Métadonnées et documentation

2.1. Introduction, définitions

2.2. Métadonnées : des standards

2.3. La documentation des données

2.4. Excursion sur sites...

- ... Entrepôts de données
- ... Portails, catalogues unifiés

2.5. Le mot de la fin

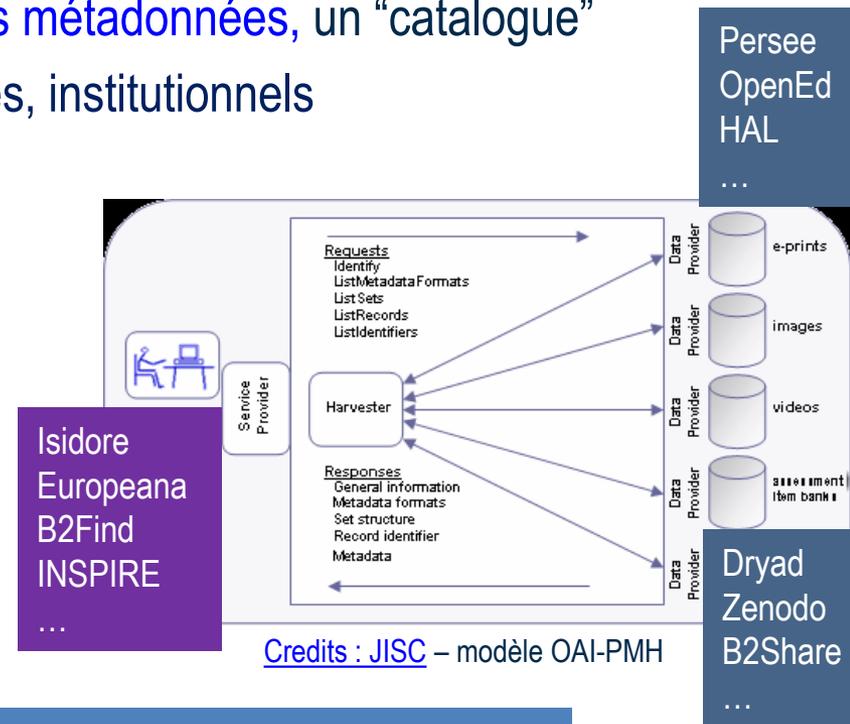


www.cnrs.fr

Des entrepôts de données et des portails

- 1 - Des sites de dépôt et diffusion de publications classiques | données Bibliothèques numériques ... Archives ouvertes ... Entrepôts de données de la recherche
 - Des documents, données publiés et des métadonnées, un “catalogue”
 - Sites locaux ou mutualisés : thématiques, institutionnels

- 2 - Des portails d'accès unifié à ces sites et à leurs données
 - Via leurs métadonnées
 - “Annuaire +”, “catalogues collectifs +”
 - Pas de ressources numériques propres
 - Valorisation, visibilité des entrepôts et données
 - Implique des métadonnées interopérables et de qualité



Catalogues d'entrepôts : re3Data

Sites de dépôt et diffusion

Accueil ENT - Catalogue Collections numérisées Bases de données Revues et livres en ligne Vidéos à la demande Guide

UNIVERSITÉ RENNES 2 COLLECTIONS NUMÉRISÉES

Outil Omeka

Consulter Collections Sur la Carte À propos

Dans les collections numérisées Recherche avancée

Contenu précédent Contenu suivant

Le Combat des Trente, poème du XIVe siècle transcrit sur le ms. original, conservé à la Bibliothèque du Roi, et accompagné de notes historiques / par Mr le chevalier de Fréminville

Poème du XIVe siècle qui relate un fameux épisode de la guerre de Succession de la Bretagne, à savoir le combat des Trente. En 1350, le baron Jean de Beaumanoir (partisan de la maison de Blois) et Robert de Bramboroug (partisan de la maison de Montfort) choisissent chacun... Voir plus

Collection: Langue, littérature et civilisation bretonne

Date : 1319 Auteur(s) : Fréminville, Christophe-Paulin de la Poix (chevalier del) Éditeur : Lafournerie et Dequier, Brest

Langue(s) : bre Sujet(s) : Trente; Combat des (1350) Bretagne (France) -- 1341-1384 (Guerre de succession)

Tags : 14e siècle Bretagne

Télécharger : 158009-10.pdf (PDF document, version 1.4 - 19.85M)

Citer ce document

Fréminville, Christophe-Paulin de la Poix (chevalier del), "Le Combat des Trente, poème du XIVe siècle transcrit sur le ms. original, conservé à la Bibliothèque du Roi, et accompagné de notes historiques / par Mr le chevalier de Fréminville." Collections numérisées - Université de Rennes 2 consulté le 24 novembre 2014. <http://bibnum.ufr-rennes2.fr/items/show/340>

Partager! Twitter Facebook StumbleUpon

Conditions d'utilisation PUBLIC DOMAIN

Institut national d'histoire de l'art INHA

Accueil Collections

Rechercher... ok

23 résultats Résultats par page : 10 Tri par : Titre (croissant)

Carnet 1 : Voyage en Nubie, 15 juin - 21 septembre 1821

Auteur Linant de Bellefonds, Louis Maurice Adolphe

Période 19e siècle

Date 1821

Sujet Linant de Bellefonds, Louis Maurice Adolphe (1799-1883) -- Archives Basse-Nubie (Égypte) -- Descriptions et voyages Dongola (Soudan ; région) -- Descriptions et voyages Nubie -- Descriptions et voyages

Cote NUM 0264 (1)

Bibliothèque d'origine Bibliothèque Centrale des Musées Nationaux

Catégorie Archives, Manuscrits et Autographes : Carnets de voyages et de notes de Maurice-Adolphe Linant de Bellefonds (1799-1883)

Carnet 10 : Divers plans de projets d'ouvrages

Auteur Linant de Bellefonds, Louis Maurice Adolphe

Période 19e siècle

Date [s.d.]

Sujet Linant de Bellefonds, Louis Maurice Adolphe (1799-1883) -- Archives

Cote NUM 0264 (10)

Bibliothèque d'origine Bibliothèque Centrale des Musées Nationaux

Catégorie Archives, Manuscrits et Autographes : Carnets de voyages et de notes de Maurice-Adolphe Linant de Bellefonds (1799-1883)

Date de publication

Affinage

Catégorie

Archives, Manuscrits et Autographes (23)

Type de document

Manuscrit (22)

Année: [dropdown]

Tous les outils gestionnaires de bibliothèques numériques et d'entrepôts de données généralistes intègrent Dublin Core

Outil Yoolib



omeka

Créer un/des “item” - étapes

Par formulaire de saisie item par item

1 – déposer des métadonnées

Dublin Core

- éléments DC simple
- éléments qualifiés - plug-in

2 – “Files” : déclaration des fichiers de données, texte, image ...

3 – Autres informations : relations ...

Par import par lot, X items

Métadonnées tableau ou XML
et données posées dans un dossier
du back-office, puis actions via
l'interface

Mêmes principes pour les autres
outils de publication numérique

Nakala et Nakalomeka, Huma-num



Gestion des données Création d'une donnée Gestion des collections Envoi d'un paquet Mon profil

Modification de la donnée ID =11280/14e95352

Fichier de données

vergy2.html

Description Générique (dcterms)

Titre: Cartulaire du prieuré de Saint-Vivant-sous-Vergy, dé
Créateur: Archives départementales de la Côte et projet CBMA
Type: flipbook
Date de création: 00/06/2010
abstract: Numérisation et mise en page d'un cartulaire manuscrit du XIVe siècle, 40 folios. Archives départementales de la Côte-d'Or / 21

Description Spécifique Nakala

Droits: accès libre
Collection: flipbooks
Format: HTML

Sauvegarder Annuler

Dublin Core
qualifié

abstract
abstract
accessRights
accrualMethod
accrualPeriodicity
accrualPolicy
alternative
audience
available
bibliographicCitation
conformsTo
contributor
coverage
created
creator
date
dateAccepted
dateCopyrighted
dateSubmitted
description
educationLevel

Nakala : plateforme de dépôt et exposition de données SHS, TGIR Huma-Num

```
<nkl:Data xmlns:nkl="http://nakala.fr/schema#"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2008/02/11/dcterms.xsd">
  <dcterms:title>Venezuela : Estado Portuguesa : Guanare : Esquina
Calle 13 et Carrera 5 : Vue d'une rue bordée d'habitations collectives
et d'une friche urbaine</dcterms:title>
  <dcterms:creator>Pouyllau, Michel</dcterms:creator>
  <dcterms:type>Image</dcterms:type>
  <dcterms:created>1983</dcterms:created>

  <dcterms:identifiser>http://medihal.archives-ouvertes.fr/medihal-
00668270</dcterms:identifiser>
  <dcterms:identifiser>http://medihal.archives-
ouvertes.fr/docs/00/66/82/70/IMG/REGARDS_Venezuela_DC065.jpg</dcterms:
identifiser>
  <dcterms:subject>[SHS:GEO] Humanities and Social
Sciences/Geography</dcterms:subject>
  <dcterms:coverage>Venezuela, Bolivarian Republic Of,
Guanare</dcterms:coverage>
  <dcterms:spatial
  xsi:type="dcterms:Point">east=-69.748947;
north=9.04549;name=Guanare</dcterms:spatial>
  <dcterms:rights>Michel Pouyllau</dcterms:rights>
  <dcterms:licence
  xsi:type="dcterms:URI">http://creativecommons.org/licenses/by-nc-
sa/2.0/fr/</dcterms:licence>
  <dcterms:source>Centre de Documentation REGARDS</dcterms:source>
  <nkl:inCollection>11280/a19a5e26</nkl:inCollection>
  <nkl:relation
```

- Import par lots
- Diffusion : Omeka
- Echanges automatisés Nakala-Omeka

OLAC, profil d'application de Dublin Core dans CoCOOn

The screenshot shows the CoCOOn website header with the CNRS logo and a search bar. Below the header is a navigation menu with 'Accueil', 'Présentation', 'Accès aux corpus', and 'Documentation'. The main content area displays the title 'Tiger and Jackal' and various metadata fields such as 'Autre(s) Titre(s)', 'Editeur(s)', 'Participant(s)', 'Description(s)', and 'Couverture spatiale'. A map of Dhaku, Nepal, is shown at the bottom left. A blue box with white text 'DC + précisions contributor et type' is overlaid on the page.

← [Affichage navigateur](#)

Notice d'une ressource

Référencement portails
 isidore
CLARIN ...

Et des identifiants pérennes ...

Identifiant OAI: oai:crdo.vjf.cnrs.fr:crdo-NEP_SYALDB_SOUND
PURL: <http://purl.org/doi/10.1017/S0022268911000088>
ARK: <http://cocoon.huma-num.fr/exist/crdo/ark:/87895/1.5-145088>
HANDLE: <http://hdl.handle.net/10670/1.cdp2wg>
Dernière modification de la notice: 2014-07-15 v1.1

DC + précisions
contributor et type

```
- <metadata>
- <olac:olac xsi:schemaLocation="http://www.language-archives.org/OLAC/1.1/ http://www.language-archives.org/OLAC/1.1/">
  <dc:title xml:lang="en">Tiger and Jackal</dc:title>
  <dc:description xml:lang="en">
    Traditional story recorded to illustrate a dialect of Nepali spoken in Achham district, Nepal. The speaker is a
  </dc:description>
  <dc:publisher>
    Laboratoire de langues et civilisations à tradition orale
  </dc:publisher>
  <dc:publisher>Centre d'Etudes Himalayennes</dc:publisher>
  <dc:contributor xsi:type="olac:role" olac:code="speaker">"DB" (anonymous)</dc:contributor>
  <dc:contributor xsi:type="olac:role" olac:code="researcher">Lecomte-Tilouine, Marie</dc:contributor>
  <dc:contributor xsi:type="olac:role" olac:code="researcher">Michailovsky, Boyd</dc:contributor>
  <dc:contributor xsi:type="olac:role" olac:code="interviewer">Lecomte-Tilouine, Marie</dc:contributor>
  <dc:contributor xsi:type="olac:role" olac:code="depositor">Michailovsky, Boyd</dc:contributor>
  <dc:contributor xsi:type="olac:role" olac:code="sponsor">Agence National de la Recherche ANR-06-CORP-030-01 (sponsor)</dc:contributor>
  <dc:format xsi:type="dcterms:IMT">audio/x-wav</dc:format>
  <dc:type xsi:type="dcterms:DCMIType">Sound</dc:type>
```

OLAC en XML exposé pour le
protocole OAI-PMH



Dryad

source/doi:10.5061/dryad.34rs7

Rechercher



About ▾ For researchers ▾ For organizations ▾ Contact us Log in Sign up



Data from: Ant and mite diversity drives toxin variation in the Little Devil poison frog

Submit data now

[How and why?](#)

Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the [Dryad Terms of Service](#). To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.  

Title	frog skin GCMS data
Downloaded	2 times
Description	Gas chromatography/mass spectrometry data from the skin of the Little Devil frog, <i>Oophaga sylvatica</i> . Frogs were collected from three populations in northwest Ecuador. Dorsal skin was isolated and alkaloids were extracted with methanol.
Download	mcguganetal_gcms_frogskin.tar.gz (82.12 Mb)
Details	View File Details

Title Ant LCMS data

Search for data

Enter keyword, DOI, etc.

Go

[Advanced search](#)

Be part of Dryad

We encourage organizations to:
[Become a member](#)
[Sponsor data publishing fees](#)
[Integrate your journal\(s\)](#), or
All of the above

[Accès au site](#)

Dryad - 2

When using this data, please cite the original publication:

McGugan JR, Bryd GD, Roland AB, Caty SN, Kabir N, Tapia EE, Trauger SA, Coloma LA, O'Connell LA (2016) Ant and mite diversity drives toxin variation in the Little Devil poison frog. *Journal of Chemical Ecology*: 1-15. <http://dx.doi.org/10.1007/s10886-016-0715-x>

Additionally, please cite the Dryad data package:

McGugan JR, Bryd GD, Roland AB, Caty SN, Kabir N, Tapia EE, Trauge O'Connell LA (2016) Data from: Ant and mite diversity drives toxin varia poison frog. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.3>

DOI	http://dx.doi.org/10.5061/dryad.34rs7
Pageviews	55
Keywords	Poison frog , alkaloid , toxin , ant , mite , mass
Date Published	2016-07-06T05:53:01Z
Scientific Names	Dendrobatidae
Spatial Coverage	Ecuador

Abstract

Poison frogs sequester chemical defenses from arthropoc

dc.relation.haspart	doi:10.5061/dryad.34rs7/1
dc.relation.haspart	doi:10.5061/dryad.34rs7/2
dc.relation.haspart	doi:10.5061/dryad.34rs7/3
dc.relation.haspart	doi:10.5061/dryad.34rs7/4
dc.relation.haspart	doi:10.5061/dryad.34rs7/5
dc.relation.isreferencedby	doi:10.1007/s10886-016-0715-x

dc.subject	mite
dc.subject	mass spectrometry
dc.title	Data from: Ant and mite diversity drives toxin variation in the Little Devil poison frog
dc.type	Article
dwc.ScientificName	Dendrobatidae
dc.contributor.correspondingAuthor	O'Connell, Lauren A.
prism.publicationName	Journal of Chemical Ecology 

Files in this package

Entrepôt Dryad

[Accès au site](#)

- [Metadata Profile](#)
 - Dublin Core qualifié + autres “espaces de noms”
 - Des référentiels pour les valeurs
 - Et des guidelines

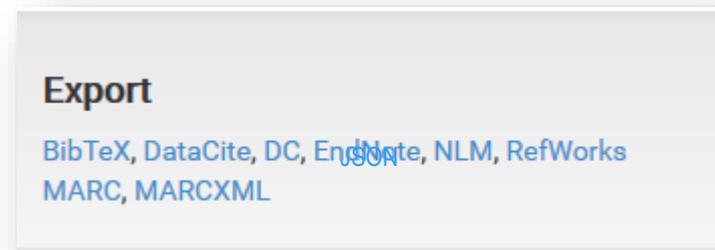
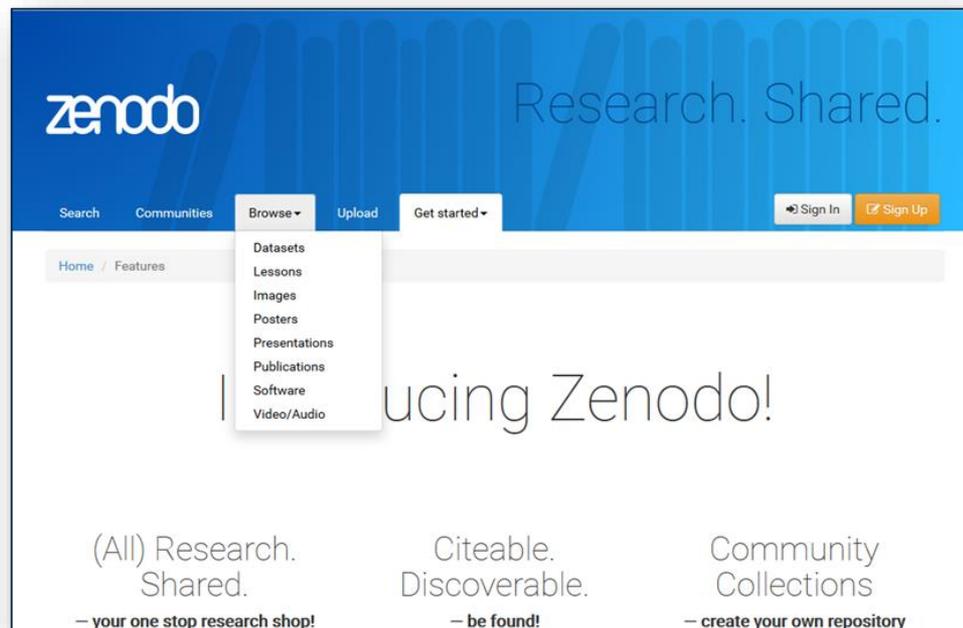
Abbreviation	Full title and Web access address
bibo	The Bibliographic Ontology: http://bibliontology.com/
dcterms	Dublin Core Metadata Initiative: http://dublincore.org/documents/dcmi-terms/
dryad	Dryad Repository: http://datadryad.org/metadata/
dwc	Darwin Core: http://rs.tdwg.org/dwc/index.htm

Abbreviation	Full title and Web access address
ITIS	Integrated Taxonomic Information System: http://www.itis.gov/ .
HIVE	Helping Integrated Vocabulary Engineering is an IMLS supported project that is developing a SKOS (Simple Knowledge Organization System) vocabulary server supporting Dryad metadata generation and user searching. (The HIVE project model is applicable beyond Dryad.) The HIVE development wiki is: https://www.nescent.org/sites/hive/Main_Page ; information on SKOS can be found: http://www.w3.org/2004/02/skos/ .
LCNAF	LC Name Authorities File, accessible as part of LC Authorities: http://authorities.loc.gov/ .
LCSH	Library of Congress Subject headings, accessible as part LC Authorities: http://authorities.loc.gov/ .
MeSH	Medical Subject Headings: http://www.nlm.nih.gov/mesh/MBrowser.html .
NBII	National Biological Information Infrastructure Biocomplexity Thesaurus: http://thesaurus.nbii.gov/portal/server.pt .
TGN	Getty Thesaurus of Geographic Names: http://www.getty.edu/research/conducting_research/vocabularies/tgn/ .
UBio	Universal Biological Indexer and Organizer: http://www.ubio.org/ .

Zenodo

[Accès au site](#)

- Publications, présentations diverses et données
 - Cadre infrastructures de recherche Europe - OpenAIRE et CERN
 - Outil Invenio
- Métadonnées
 - Internes : profil d'application de **MaRC**, modèle OpenAIREplus
 - Dépôt diapo suivante
 - Exports variés
 - + entrepôt OAI-PMH



Zenodo, dépôt par formulaire

Delete Save Submit ...

New upload

Instructions: (i) Press "Save" to save your upload for editing later, as many times you like. (ii) Upload and remove extra files in the bottom of the form. (iii) When ready, press "Submit" to finalize and make your upload public.

Type of file(s)

 required ▼

Publication Poster Presentation Dataset Image Video/Audio
Software Lesson

Basic information

 required ▼

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.

Pre-reserve 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.

License

 required ▼

Access right * Open Access
 Embargoed Access
 Restricted Access
 Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

License *

Required. The selected license applies to all of your files displayed in the bottom of the form. If you want to upload some files under a different license, please do so in two separate uploads. If you think a license missing is in the list, please inform us at info@zenodo.org.

Communities

 recommended ▼

Any user can create a community collection on Zenodo ([browse communities](#)). Specify communities which you wish your upload to appear in. The owner of the community will be notified, and can either accept or reject your request.

Communities

Funding

 recommended ▼

Zenodo is integrated into reporting lines for research funded by the European Commission via OpenAIRE (<http://www.openaire.eu>). Specify grants which have funded your research, and we will let your funding agency know!

Grants

Optional. Note, a human Zenodo curator will validate your upload before reporting it to OpenAIRE, and you may thus experience a delay before your upload is available in OpenAIRE.

... Un pas de plus : dépôt par API ou Web service ?

- Echanges entre machines, web ici “hors formulaire”
 - Protocole du web, HTTP + outils spécifiques, ex cURL - ou piloté par une page d'interface plus humaine
 - Gain de temps, sécurité-fiabilité
 - Connaître et pratiquer l'interface, la syntaxe de requête
 - Préparer les fichiers de données et métadonnées selon spécifications
- Zenodo : [dépôt par API “REST”](#)
 - Métadonnées : des éléments “verbalisés”, syntaxe JSON

<code>publication_date</code>	String	Yes	Date of publication in ISO8601 format (<code>YYYY-MM-DD</code>). Defaults to current date.
<code>title</code>	String	Yes	Title of deposition.
<code>creators</code>	Array of objects	Yes	The creators/authors of the deposition. Each array element is an object with the attributes: <ul style="list-style-type: none">• <code>name</code> : Name of creator in the format <i>Family name, Given names</i>• <code>affiliation</code> : Affiliation of creator (optional).• <code>orcid</code> : ORCID identifier of creator (optional).• <code>gnd</code> : GND identifier of creator (optional). Example: <pre>[{'name': 'Doe, John', 'affiliation': 'Zenodo'}, { 'name': 'Smith, Jane', 'affiliation': 'Zenodo', 'orcid': '0000-0002-1694-233X'}, { 'name': 'Kowalski, Jack', 'affiliation': 'Zenodo', 'gnd': '170118215'}]</pre>

- HAL, [dépôt par API “SWORD”](#) : métadonnées XML-TEI + référentiels AureHAL

beQuali et DDI

Banque d'enquêtes qualitatives en sciences sociales

L'enquête en bref :

accueil **les enquêtes** principes outils

← L'enquête en bref Contents de l'enquête

L'Europe saisie par les rôles parlementaires
Olivier Rozenberg - France, 1999-2005

Général

DDI Id edsp_bequali_sp3
Nom Le Parlement français et l'Union européenne
Résumé Enquête par entretiens individuels semi-directifs réalisée avec des parlementaires sur la mobilisation en matière européenne
Discipline Science politique
Mots clés Parlement, parlementaires, institutions politiques, Europe, représentation
Classification Sociologie des institutions politiques
Pays France
Période couverte 1993-2005

Intervenants

Auteur(s) Olivier Rozenberg

Univers

Niveau de comparatisme	National
Couverture géographique	Non pertinent
Unités spatialisées	Assemblée nationale, Sénat, Gouvernement, Parlement Européen
Unités d'observation	Individus
Groupes de population ciblés	Parlementaires
Dates de l'enquête	1999-2005

Méthode

Dates de récolte des données	1999-2003
Périodicité	Unique (quelques témoins ont été interviewés deux fois)
Accès aux observations	Recrutement par sollicitation directe des parlementaires
Echantillonnage	Diversification de l'échantillon selon l'expérience en matière européenne
Segmentation	Segmentation selon la fonction institutionnelle occupée
Technique de collecte	Entretiens semi-directifs

Corpus

Nombre de documents	
Types de documents	

Général

Intervenants

Univers

Méthode

Corpus

Analyse

Edition

Un instrument de recherche détaillé est disponible en téléchargement. Il présente l'ensemble des documents archivés, mis à disposition ou non.

Menu : accès aux différentes sections de métadonnées

```
<ddi:geogCover>PACA</ddi:geogCover>
<ddi:geogCover>Auvergne</ddi:geogCover>
<ddi:anlyUnit>Individus</ddi:anlyUnit>
<ddi:universe level="study">Niveau d'observation:</ddi:universe>
<ddi:universe level="study">national</ddi:universe>
<ddi:universe level="study">Groupes de population ciblés:</ddi:universe>
<ddi:universe level="study">citoyens ordinaires</ddi:universe>
<ddi:collDate date="1978-01-09" event="start">9 January 1978</ddi:collDate>
<ddi:collDate date="1978-03-17" event="end">17 March 1978</ddi:collDate>
<ddi:dataKind type="text">Documents textuels</ddi:dataKind>
<ddi:dataKind>Transcriptions fines avec indication du paraverbal</ddi:dataKind>
</ddi:sumDscr>
</ddi:studyInfo>
<ddi:method>
<ddi:dataColl>
<ddi:timeMeth>Entretiens uniques</ddi:timeMeth>
<ddi:collMode>Enquête par entretiens</ddi:collMode>
<ddi:sampProc>La sélection des personnes interviewées a été faite par les enquêtés
<ddi:resInstru>consigne (entretien non directif)</ddi:resInstru>
<ddi:collSitu>La même question a été posée à tous les enquêtés : "Voulez-vous qu'il y ait un
</ddi:dataColl>
</ddi:method>
<ddi:dataAcqs>
<ddi:setAvail>
<ddi:fileQnty>266 fichiers</ddi:fileQnty>
```

Source : présentation interne, décembre 2014

Data Documentation Initiative (DDI)



Source: <http://www.ddialliance.org/>

Un standard de description des enquêtes et études de terrain,
de leur conception aux résultats et à leur diffusion

Des outils de création

[Source](#)

Metadata provided by
GESIS ZACAT European
Value Study
gesis
Leibniz Institute
for the Social Sciences
Go to Nesstar Server

- Study Information
- Overview
- Technical Information
 - Sampling
 - Data Collection
- Datasets
 - Access Policy
 - Data files
 - ZA4797_v1-1-0
 - Variable Groups
 - Archive and ID Variables
 - Weight [ZA4797]
 - Perceptions of Life
 - Politics and Society
 - Work
 - Religion and Morale
 - Family
 - National Identity
 - Environment
 - Demographics and Interview Characteristics
 - Additional country-specific variable
 - Variable Search

<ddi> Metadata in XML

Contenu de l'enquête : le tri des documents

Filtres par axe de classement :

Filtre par localisation :

DDI + Dublin Core, METS, EAD
Saisies une fois → conversions → usages et suite du cycle de vie

Interopérabilité avec les Archives de France, le portail Quetelet ... Isidore

Quand des Français, des Anglais et des Belges (francophones) parlent d'Europe
 CDSP | cdsp_bequali_sp1
 Sophie Duchesne, Elizabeth Frazer, Guillaume Garcia, André-Paul Frogner, Florence Haegel, Virginie Van Ingelgom

date de production : 2006

date(s) de collecte : 2005-2006

thèmes : science politique, sociologie politique, études européennes

résumé : Enquête par entretiens collectifs sur la politisation et la polarisation des réactions à l'égard de l'intégration européenne

description détaillée (XML-DDI) commander le(s) fichier(s) haut de page



KnB – Knowledge Network for Biocomplexity

knB

ABOUT

DATA

SHARE

TOOLS

Jump to:

DOI or ID

Go

SIGN IN

Data Table

Entity Name: Definitions-PBB-Data-2009.csv

Object Name: Definitions-PBB-Data-2009.csv

Online Distribution Info: ecogrid://knB/pbb.20.15

Size: 4773 byte

Text Format

Number of Header Lines

1

Record Delimiter

Attribute Orientation

Simple Text

Field Delimiter

Number Of Records: 52

Attribute Information

Variables

Plant Group ...

Plant Group ...

Phenophase ...

Phenophase ...

Phenophase ...

Phenophase ...

Site State ID

Site State De...

Name: Plant Group Code

Label

Definition: Project BudBurst defined Deciduous Trees and Shrubs; Grasses; Conifers; Evergreen

Storage Type: VarChar2 (128 Char)

Measurement Type: nominal

Measurement Domain: Enumerated Domain (Data)

EML
Ecological Metadata language



Sampling, Processing and Quality Control Methods

Methods

Step 1

Description

Regular Reports

Regular Reports: Event-based monitoring where participants watch a specific plant throughout the year and record the date on which specific phenophases occur.

Participants watch their plant over time and record on their Regular Report datasheets the date when specific phenophases first occur.

If a participant misses the date a phase occurs, they skip that phase and move on to the next one.

Regular Report phenophases that participants watch for are dependent upon the plant group and include the following:

Deciduous Trees and Shrubs

Bud Burst

Date when the protective scale coating is shed from the bud exposing tender new growth tissues of one or more flower buds or leaves

First Leaf

ISIDORE

Portail d'accès unifié aux publications
et données de recherche en SHS

Métadonnées → Accès collections et
objets sur les sites source (plus de 3000)

Moissonnage
OAI-PMH :
Dublin Core et
quelques
autres, RSS ...

Curation :
indexation par
thésauri

Réexposition
en RDF →
web de
données,
DARIAH ...

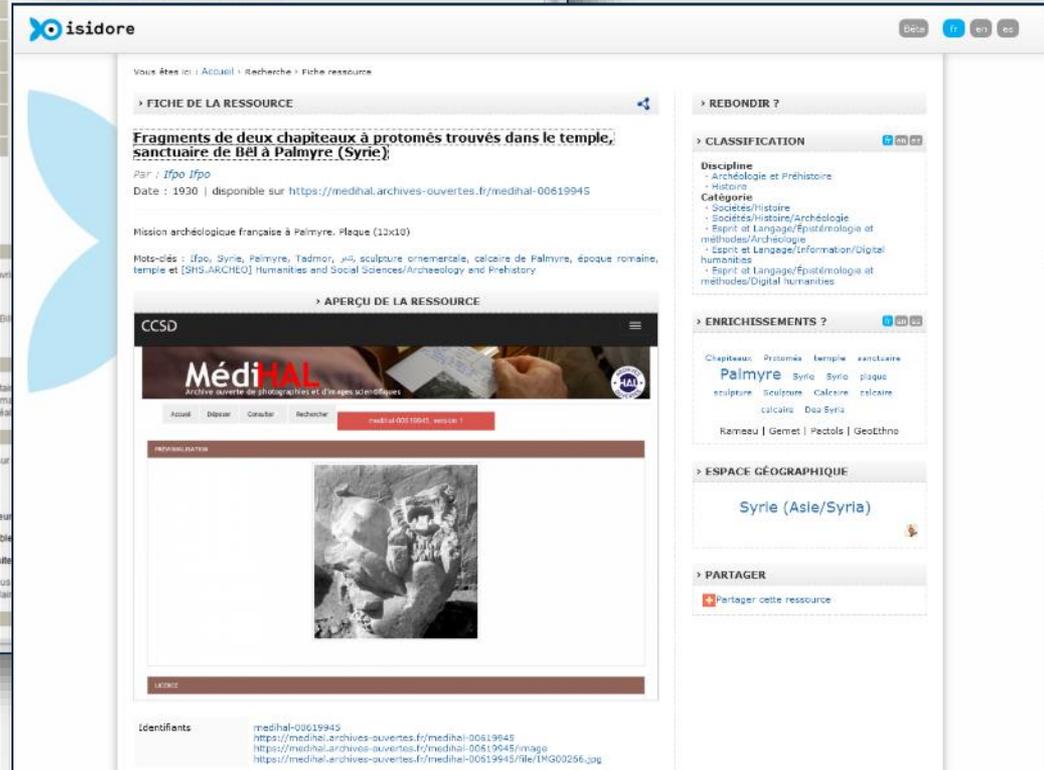


Tableau 1: Métadonnées pour les séries de données géographiques et les ensembles de séries de données géographiques

Référence	Éléments de métadonnée	Cardinalité	Condition
1.1	Intitulé de la ressource	1	
1.2	Résumé de la ressource	1	
1.3	Type de ressource	1	
1.4	Localisateur de la ressource	0..*	Obligatoire s'il existe un URL permettant d'obtenir davantage d'informations sur la ressource et/ou les services liés à l'accès.
1.5	Identificateur de ressource unique	1..*	
1.7	Langue de la ressource	0..*	Obligatoire si la ressource inclut des informations textuelles.
2.1	Catégorie thématique	1..*	
3	Mot-clé	1..*	
4.1	Rectangle de délimitation géographique	1..*	
5	Référence temporelle	1..*	
6.1	Généalogie	1	
6.2	Résolution spatiale	0..*	Obligatoire pour les séries de données et les ensembles de séries de données pour lesquels une échelle équivalente ou une distance de résolution peuvent être indiquées.
7	Conformité	1..*	
8.1	Conditions applicables à l'accès et à l'utilisation	1..*	
8.2	Restrictions concernant l'accès public	1..*	

Projet INSPIRE

- Métadonnées : profil d'application de ISO 19115
 - Nbx domaines : importance de l'information géospatiale
 - Vers des portails thématiques et nationaux,
 - Vers le Portail européen INSPIRE
 - Données administrations publiques (directive) et scientifiques (? ...)

Welcome to the INSPIRE geoportal

The INSPIRE Directive requires the Commission to establish a community geo-portal and the Member States shall provide access to their infrastructures through the geo-portal as well as through any access points they themselves decide to operate.

More...

Discovery / Viewer

Search, discover and access geographic information provided by European governmental, commercial, and non-commercial organizations.

More ...



Validator

The purpose of the INSPIRE Metadata Validator is to test the compliance of INSPIRE metadata with the INSPIRE Metadata Regulation.

More ...



Metadata Editor

Create metadata according to the INSPIRE implementing rules.

More ...



Resource Browser

Technical insight into resource metadata

More ...



- CArGOS, catalogue de données géographiques en SHS
 - Exemple
 - Mots-clés thésaurus INSPIRE-GEMET Spatial Data Themes
- Portail des données marines (Ifremer)
- Indigeo, Infrastructure géographique de données géospatialisée ssur l'environnement, France Ouest
- INSPIRE geoportal : catalogue, outils

DataOne, portail outillé

- Data Observation Network for Earth et ses membres : KNB ([Knowledge Network for Biocomplexity](#)), USGS ...
 - Terre, environnement, biodiversité
 - Soutenir le partage des données et les bonnes pratiques
- Métadonnées [EML](#), [FGDC](#) (ISO 19115), DC et divers
- Développement et recommandation d'outils de mise en œuvre variés
 - Créer et gérer des entrepôts de données et métadonnées : [MetaCat](#)
 - ... insistance sur métadonnées et documentation

Meta Cat NCEAS Metacat: Metadata and Data Management Server

Go INDEX NEXT

Metacat Administrator's Guide

Metacat is a repository for data and metadata (documentation about data) that helps scientists find, understand and effectively use data sets they manage or that have been created by others. Thousands of data sets are currently documented in a standardized way and stored in Metacat systems, providing the scientific community with a broad range of science data that—because the data are well and consistently described—can be easily searched, compared, merged, or used in other ways.

- Download Metacat

Version: 2.4.1

DataONE : outils

- Portail fédérateur “avancé” : outil [Mercury](#)
 - Moissonnage par OAI-PMH, Z39.50, OpenSearch de divers entrepôts
 - **Métadonnées XML (DC, EML, Darwin Core, FGDC, ISO 19115)** et “données clé”
 - interface web : facettes, cartes, exports ... et accès aux données dans leur contexte
- ... Mais aussi : liste d’outils de manipulation “locale” de **métadonnées**, indépendante d’un système de diffusion, pour divers usages
 - [Morpho](#) : métadonnées EML
 - [MetadataParser](#) : métadonnées FGDC
 - Plus d’info sur le [site USGS](#), [plus encore](#) sur le site FGDC
 - CatMDedit : Dublin Core
 - ... Editeurs et bases de données XML, outils images





2 – Métadonnées et documentation

2.1. Introduction, définitions

2.2. Métadonnées : des standards

2.3. La documentation des données

2.4. Excursion sur sites...

2.5. Le mot de la fin : en pratique...



www.cnrs.fr

En pratique ...

- “Record all the information needed for you and others to understand and use the data in the future” ... [EUDAT, introduction to Metadata](#)

Check-list typologique : contenu, structure, méthodologie scientifique, traitements et caractéristiques techniques/informatiques, propriété intellectuelle et droits/limites d'utilisation ... complémentarité métadonnées et documentation

- Une équipe multi-compétences
- Commence dès le processus de collecte-crétion-organisation des données
 - Le meilleur moment pour la qualité et la dynamique ...
- Vers de l'information utile, échangeable et pérenne → standards
 - Partiellement automatisable ?
 - Métadonnées extraite des données, de la documentation puis convertie, enrichie quand nécessaire



En pratique - 2

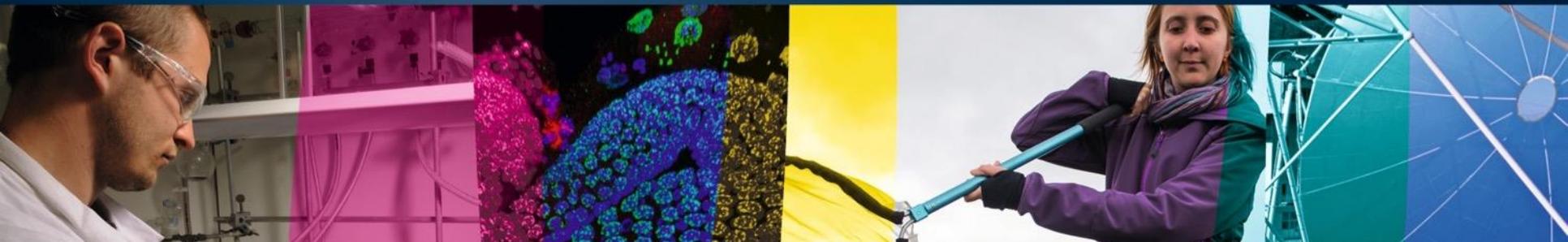
- Utiliser des vocabulaires d'éléments standards
 - Intérêt premier : basé sur un modèle réfléchi et mûri - puis permet l'interopérabilité
 - La base de données locale peut l'implémenter différemment ...
- Et renseigner les valeurs de manière homogène
 - Valeur ajoutée importante
- Des standards, des spécifications, des outils
 - Standards reconnus par la communauté, la thématique
 - Standards selon objectifs, applications cible
 - Standards liés au type de données
- Les exprimer dans une ou plusieurs syntaxes standard actuelles : XML, JSON
 - ... Selon flux de données prévus
 - Dépôt par lot dans l'entrepôt choisi, archivage pérenne ... ?
 - Penser interopérabilité technique d'un entrepôt local
 - Potentiellement évolutives
 - Quand on en "tient" une et qu'on travaille en équipe multi-compétences, on tient les autres ...





www.cnrs.fr

Merci pour votre attention





www.cnrs.fr

Références

Références - 1

- 1 - [Data Management Plan for the PhD project: Development and Application of a Monitoring System to Assess the Impacts of Climate and Land Cover Changes on Eco-Hydrological Processes in an Eastern Andes Catchment Area](#)
- 2 - JISC, Digital Media, [Guides A to Z](#)
- 3 - UK Archives, Create & Manage data – [Organising data](#)
- 4 - Université d'Edinburg, exemple : [Naming Conventions](#)
- 5 - Université de Lausanne, UNIRIS-UNIL, [gestion des données de recherche, organiser et stocker](#), ou : [Règles de nommage des documents électroniques](#),
- 6 - CINES, [Expertise Formats](#) et outil [FACILE](#)
 - [non publié] Alexandre Granier (CINES), Quels formats de fichiers pour l'interopérabilité ?, ANF Humanum "Gérer les données de la recherche : de la création à l'interopérabilité", 21-25 septembre 2015
- 7 - Groupe PIN, "Préservation des informations numériques", [Formats](#)
- 8 - Questions ENSSIB, [règles de nommage des fichiers et répertoires](#)
- 9 – Huma-num, [Guides de bonnes pratiques](#)

Références - 2

- 10 - M.-C. Jacquemot-Perbal et F. Cosserat / CNRS-Inist, [Gestion et diffusion des données de la recherche](#), journée de formation, URFIST de Rennes (mai 2015)
- 11 - CNRS-Inist, [“Le libre accès aux résultats de la recherche dans le cadre d’Horizon 2020”](#), 5 Tutoriels, 2015
- 12 - RDA, Research data Alliance, [Outputs & Recommendations](#)
Recommandations et réalisations des Groupes de travail (WG) : Metadata Standards Directory, Data Description Registry Interoperability ...
- 13 - Van den Eyden V, Corti L, Woollard B et al (2011). [Managing and sharing data. Best practice for researchers](#)
- 14 - Sarah Jones (DCC), [Managing research data and Horizon 2020](#), Consorcio Madroño conference on Data Management Plans and Horizon 2020, ETSI Industriales, Madrid, 25th February 2015
- 15 - BioSharing : <http://www.biosharing.org>, notamment [Standards](#)
- 16 - Knowledge Network for Biocomplexity (KNB) et Data Observation Network for Earth, [dataOne](#) : Tutoriels, outils ...
- 17 - [EUDAT, introduction to Metadata](#)
- 18 – FOSTER, [Documentation, Organization and Storage](#)

Références - 3

- 18 - Université Paris-Diderot, Paris-Descartes, [Réaliser un plan de gestion de données](#), V1, 2015-01
- 19 - Digital Curation Centre, [Checklist for a Data Management Plan](#), 2015
- 20 - [Data management Planning Tool](#), DMPTool, [Data Management General Guidance](#)

- 21 - Commission européenne, Direction Générale de la Recherche et de l'Innovation, [Lignes directrices pour la gestion des données dans Horizon 2020](#), V2.1, 15-02-2016 – traduction française
- 22 - Commission européenne, Direction Générale de la Recherche et de l'Innovation [Lignes directrices pour le libre accès aux publications scientifiques et aux données de la recherche dans Horizon 2020](#), V2.1, 15-02-2016 – traduction française
- 23 - Commission européenne, [European Open Science Cloud](#), [A Cloud on the 2020 Horizon. Realising the European Open Science Cloud: first report and recommendations](#), 20 juin 2016