


Atelier « Carnets de terrain électroniques », Montpellier, 28-29/03/2018

Sécuriser les données produites par les carnets de terrain électroniques

Marie-Claude Quidoz (CNRS/CEFE)





Ce(tte) œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.

Vous êtes autorisé à :

Partager — copier, distribuer et communiquer le matériel par tous moyens et sous tous formats

Adapter — remixer, transformer et créer à partir du matériel

Selon les conditions suivantes :



Attribution — Vous devez mentionner le nom de l'auteur de la manière suivante :
« Marie-Claude Quido, CEFE-CNRS, 2018 »



Pas d'Utilisation Commerciale — Vous n'êtes pas autorisé à faire un usage commercial de cette Oeuvre, tout ou partie du matériel la composant.



Partage dans les Mêmes Conditions — Si vous modifiez, transformez ou adaptez cette œuvre, vous n'avez le droit de distribuer votre création que sous une licence identique ou similaire à celle-ci.

Voir la version intégrale de la licence : <http://creativecommons.org/licenses/by-nc-sa/4.0/>



Plan

- Des exemples de mauvaises sécurisation
- Risques afférents aux fichiers
- Pistes d'améliorations possibles
- Risques spécifiques aux carnets de terrain électronique
- Pistes d'améliorations possibles



Problème de logiciel

Sujet : [mate-shs] Problème conversion de données

Date : Fri, 9 Feb 2018

Bonjour à toutes et à tous,

Y aurait-il parmi vous quelqu'un qui utilise Modalisa?

J'ai travaillé avec la version 4.0 en 2008 et sauvegardé des données en un format qui n'est pas compatible avec les autres logiciels, et avec R en particulier...

Il existait bien auparavant un package ("rgrs") qui permettait cette compatibilité mais son remplacement par "questionr" ne permet plus le traitement des fichiers de données de sauvegarde .TXT issus de Modalisa

Comme nous n'avons pas Modalisa au labo ni à l'université, j'aimerais savoir si une âme charitable ayant accès au logiciel accepterait de me convertir ce fichier de données en .csv ?

Bien à vous,

Problème matériel

- En 1984, en collaboration avec le CNUSC, sauvegarde d'une partie de l'Ecothèque Méditerranéenne sur bande



1. BANDE MAGNETIQUE N°1 : DONNEES ORNITHOLOGIQUES

- . F 101/VENT/VENTOUX, BLONDEL BADAN
325 relevés, CMR/CDR/CDE, 32 variables
- . F 102/CORS/CORSE, BLONDEL
262 relevés, CMR/CDR/CDE, 32 variables
- . F 103/ALPE/ALPILLES, BLONDEL
323 relevés, CMR/CDR/CDE, 32 variables
- . F 104/DURA/DURANCE, BLONDEL GALLARDO

- 25 ans après, intérêt et souhait de réutilisation des données avec perspectives de répétitions
- Mais impossibilité de trouver un lecteur de bande
(CINES, CNES, INIST, CC-IN2P3, INSU, DSI, IDRIS, IBM, Musée de l'informatique, ACONIT,)

Problème de compréhension (1)

- Support papier

```

7EUD
/ADV2/ADVENTICES DU MONTPELLIERAIS MAILLET ENSAM PASSAGE 2
292 RELEVES 52 VARIABLES NOTE CDF= 1 NOTE CDE= 1
VOIR CARTES GUIDES DU FICHER ADVI
1 0 474 44 3 16 05 2 2 01 01 1 16 1 06 14 7 6 1 0 00 10 00 1 4 2 31 3 313 2 1 1
4 4 1 4 03 01 1 10 5 1 120018530 4 07 78 07780 5 4 00 1 4 2 31 3 313 2 1 1
1 0 574 44 3 16 05 2 2 01 01 1 16 1 06 14 7 6 1 0 00 10 00 1 4 2 31 3 313 2 1 1
4 4 1 3 03 00 1 10 5 3 120006630 3 07 78 07780 3 4 00 1 4 2 31 3 313 2 1 1
2 0 682 46 3 16 04 2 2 06 03 1 15 1 03 10 7 2 1 1 14 24 00 1 2 2 43 3 433 3 3 1
4 4 1 3 04 02 3 01 1 3 120013370 3 07 78 07780 4 4 00 1 4 2 21 3 213 2 1 3
3 1 786 45 3 16 04 2 1 06 03 1 15 1 04 08 8 3 1 1 24 00 00 1 4 2 21 3 213 2 1 3
4 4 1 4 04 04 3 01 2 1 038324540 3 07 78 07780 5 4 00 1 4 2 21 3 213 4 2 3
2 0 831 48 2 02 04 4 3 02 01 1 16 1 12 10 8 3 1 0 20 00 00 1 4 2 21 3 213 4 2 3
4 4 1 4 02 05 3 01 2 3 157112000 3 07 78 07780 5 4 00 1 4 2 21 1 211 0 0 0
0 0 956 76 6 37 11 4 3 02 01 1 16 1 03 14 7 2 1 0 10 10 00 1 4 2 21 1 211 0 0 0
4 4 1 4 02 01 1 10 5 1 442815550 4 07 78 07780 5 4
  
```

- Documentation associée

Colonnes

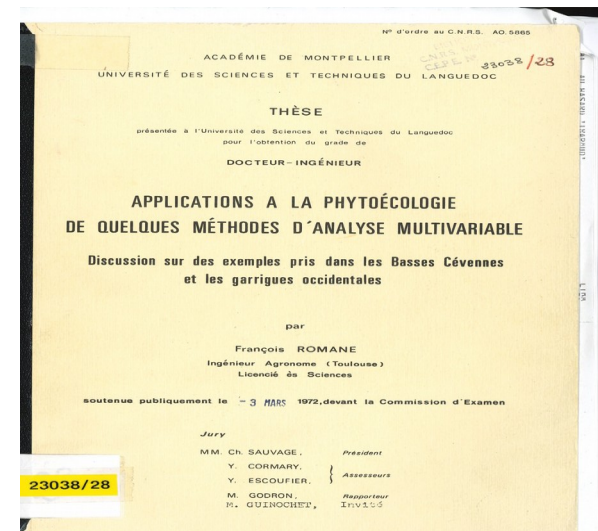
	1	Numéro de carte CMR	M -	31	Formation végétale
A -	2 à 4	Numéro de relevé	N -	32-35	1ère espèce dominante
B -	5	Géomorphologie	D -	36-39	2ème espèce dominante
C -	6-7	Faciès de végétation	P -	40-41	Altitude
D -	8	Répétition des relevés	Q -	42	Exposition

Problème de compréhension (2)

- Pas la bonne documentation : elle avait été écrite pour la retranscription sur cartes magnétiques

REGLES DE TRANSCRIPTION SUR CARTES PERFOREES DES OBSERVATIONS
ECOLOGIQUES ET PHYTOSOCIOLOGIQUES CONTENUES DANS UN ENSEMBLE
DE RELEVES

- La solution trouvée : utiliser les annexes des articles / thèses rédigés utilisant le jeu de données





Principaux risques sur un fichier

- Existence de 4 risques menaçant inéluctablement un fichier sur une longue période :
 - ✓ Obsolescence matérielle
 - ✓ Obsolescence logicielle
 - ✓ Obsolescence du format de fichier
 - ✓ Perte de la signification du contenu
- Et d'autres :
 - ✓ Mauvais étiquetage (du lieu de rangement, du support, ..)
 - ✓ Manque de rigueur dans le nom des fichiers
 - ✓



Les défis, orientations et choix pour l'archivage électronique

Voici un document que j'ai créé en 1998...

De quoi s'agit-il déjà ?
Est-ce bien ce qui est
indiqué sur la
disquette ?

METADONNEES DESCRIPTIVES + IDENTIFICATION
UNIQUE et PERENNE

La disquette est-elle
toujours en bon état ?

SUPPORT : VEILLE +
MIGRATION PHYSIQUE



Mon portable, acheté
en 2013, n'a pas de
lecteur de disquette...

ENVIRONNEMENT MATERIEL :
VEILLE TECHNO et ANTICIPATION

Ça marche ! Mais j'ai
perdu toute ma mise
en forme...

INTEGRITE AUTHENTICITE

J'ai trouvé le logiciel,
mais puis-je l'installer
et l'utiliser sous
Windows 8 ?

SYSTEME D'EXPLOITATION

J'ai créé ce document avec Claris
Works. Comment retrouver ce
logiciel ? Quel est le format du
document ?

ENVIRONNEMENT LOGICIEL : privilégier
les FORMATS DURABLES + MIGRATION
LOGIQUE



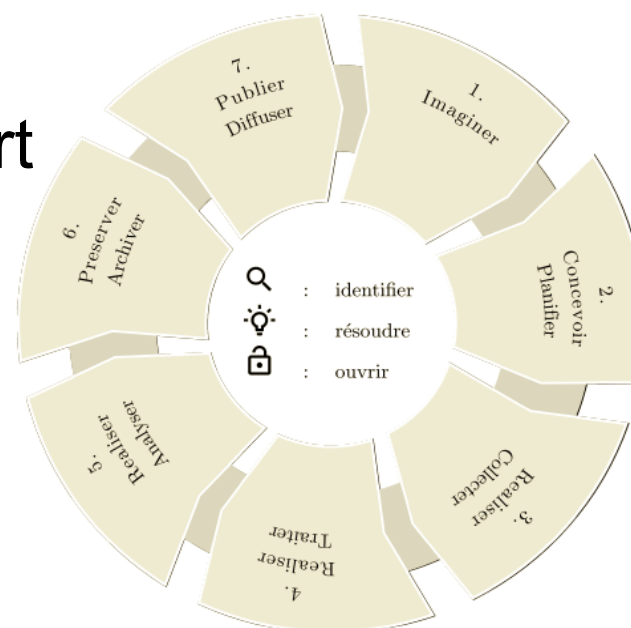
Pistes d'améliorations possibles


- Inclure cette réflexion dès le départ

- ✓ Cycle de vie de la donnée
- ✓ Plan de gestion des données

- Prévoir :

- ✓ Du temps
 - ▶ Veille technologique / Documentation / Migration
- ✓ Un budget
 - ▶ Personnel / Matériel / Logiciel / ...






Pistes d'améliorations possibles pour éviter les problèmes matériels

- Dupliquez les sauvegardes
- Testez régulièrement les supports
- Conservez les supports dans des lieux sains et sécurisés
- Suivez les évolutions technologiques
- Migrez les fichiers sur un autre support si nécessaire



Pistes d'améliorations possibles pour éviter la perte de sens

- Documentez les attributs (colonnes) du fichier
 - ✓ Métadonnées
- Soyez méthodique dans les appellations
 - ✓ Fichiers / répertoires / attributs
- Stockez tous les éléments au même endroit
 - ✓ Faites du ménage dans tous vos fichiers



Pistes d'améliorations possibles pour éviter les problèmes logiciels

- Privilégiez un logiciel
 - ✓ Largement utilisé
 - ✓ Avec les sources publiées
 - ✓ Offrant un format d'enregistrement durable
- Suivez les évolutions technologiques
- Migrez les fichiers sur une nouvelle version (ou un nouveau) logiciel si nécessaire



Pistes d'améliorations possibles pour éviter les problèmes de format

- Privilégiez des formats
 - ✓ Publiés (spécification existante, accessible et libre de droit)
 - ✓ Largement utilisés (ou promis de l'être)
 - ✓ Normalisé (ou représentant un standart de fait)
- Testez votre fichier avec l'outil du CINES
 - ✓ <https://facile.cines.fr/>
- Migrez les fichiers sur un nouveau format si nécessaire



Formats conseillés

- Formats durables (cf liste faite par le CINES) :
 - ✓ PDF
 - ✓ CSV / TXT
 - ✓ TIFF / GeoTIFF / JPEG* / PNG
 - ✓ MPEG / FLAC
- Format bureautique ouvert :
 - ✓ ODS / ODT
 - ✓ XLSX / DOCX



Un conseil important

- Ne pensez pas uniquement format ouvert mais pensez aussi format durable pour une sauvegarde à long terme (< 15 ans)
- Dans le monde des tableurs, où à ce jour les seuls formats durable sont CSV / TXT, n'oubliez pas de prendre en compte, lors de la création de vos fichiers, la perte de fonctionnalités de ce format par rapport aux formats natifs

Exemples

	A	B	C	D	E	F
1	numero_bague	personne	date	etang	type_bague	
2	numéro de la bague métallique ou darvic. Ne pas rentrer d'information incomplète	code de la personne qui a perdu la bague. Laisser le champ vide si l'information manque	Format : JJ/MM/AAAA	Entrer le code de l'étang	Important : Entrer si c'est une bague darvic ou métal qui est perdue.	
3						
4						
5						
6						
7						
8						

Fichier « bague » en ods

	A	B	C	D	E
1	numero_bague	personne	date	etang	type_bague
2					


Le même fichier en csv (perte des commentaires)

	A	B	C	D	E
1	jour	annee	date		
2	10	2016	11/03/16		
3	11	2016	12/03/16		

Fichier « date » en ods

	A	B	C	D
1	jour	annee	date	
2	10	2016	11/03/16	
3	11	2016	12/03/16	

Le même fichier en csv (perte de la formule de calcul)



Risques spécifiques aux carnets de terrain électroniques

- Plus de support papier
 - ✓ Moyen de sauvegarde couramment utilisé (photo des pages du carnet de terrain) n'est plus possible
- Monde uniquement numérique
 - ✓ Les 4 risques cités précédemment sur un fichier si vous enregistrez vos données sur fichier ou sur un SGBD local
- Panne informatique / vol de matériel / ...



Pistes d'améliorations possibles

- Sur le terrain
 - ✓ Faites un enregistrement vocal en parallèle
 - ✓ Ne partez jamais sans un carnet papier
 - ✓ Si vous n'êtes pas dans une zone connectée, sauvegardez chaque soir en format texte vos données saisies sur un support externe
 - ✓ Si vous êtes dans une zone connectée, privilégiez un carnet de terrain qui alimente directement le système d'information de votre structure



Pistes d'améliorations possibles

- En fin de campagne
 - ✓ Stockez vos données sur le système d'information de votre structure qui se chargera des risques d'obsolescences (matériel, logiciel et format)
 - ✓ Documenter vos données (métadonnées) et la base de données (Modèle Conceptuel de Données)



Bibliographie

- Journée de sensibilisation à la sécurisation et à la pérennisation des données, rBDD, Novembre 2014
- <http://rddb.cnrs.fr/spip.php?article172>
- Groupe de travail “Pérennisation de l'Information Numérique”, association Aristode
<http://pin.association-aristote.fr/doku.php>
- Guide méthodologique : le format de fichier PDF, CINES & NUMEN, Août 2012
- <https://www.huma-num.fr/sites/default/files/guide-format-fichiers-pdf.pdf>
- Guide méthodologique pour le choix de formats numériques pérennes dans un contexte de données orales et visuelles, CINES & Huma-Num, Avril 2011
- <https://www.huma-num.fr/sites/default/files/guide-formats-numeriques.pdf>
- Guide méthodologique pour l'archivage des bases de données, CINES, Avril 2013
https://alfresco.cines.fr/alfresco/d/d/workspace/SpacesStore/ca323bcc-804a-43e3-822f-5d18fa304dff/GM_archivage_BDD.pdf